GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION
SPONSORED PROJECT INITIATION

aLG

Date: 10/14/78

Project Title: Incentives, Organizational Design and Iterative Communication

Project No: M-50-627 Green card

Co- Project Directors: Susan I. Cohen and Matthew J. Sobel

Sponsor: National Science Foundation

Agreement Period: From 7/1/78 Until 12/31/80

Type Agreement: Grant No. SOC78-05770

Amount: $45,823 NSF
8,106 GIT (M-50-318)
$53,929

Reports Required: Annual Progress Report; Final Project Report

Sponsor Contact Person (s):

Technical Matters
(Program Officer)
James H. Blackman
Economics Program
Electrical Sciences and Analysis Section
Division of Engineering
National Science Foundation
Washington, D. C. 20550
(202) 632-5968

Contractual Matters
(thru OCA)
(Grants Manager)
Ms. Mary Frances O'Connell
Grants Specialist
MPE/BBS/SE Branch
Division of Grants & Contracts
National Science Foundation
Washington, D. C. 20550
(202) 632-2858

Defense Priority Rating: n/a

Assigned to: Industrial Management (School/Laboratory)

COPIES TO:

Project Director
Division Chief (EES)
School/Laboratory Director
Dean/Director—EES
Accounting Office
Procurement Office
Security Coordinator (OCA)
Reports Coordinator (OCA)

Library, Technical Reports Section
EES Information Office
EES Reports & Procedures
Project File (OCA)
Project Code (GTRI)
Other

CA—3 (3/76)

GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION

Date: 4/23/81

Project Title: Incentives, Organizational Design and Iterative Communication

Project No: M-50-627

Project Director: Dr. M. J. Sobel

Sponsor: NSF

Effective Termination Date: 12/31/80

Clearance of Accounting Charges: 12/31/80

Grant/Contract Closeout Actions Remaining:

- _ Final Invoice and Closing Documents
- x Final Fiscal ~~Report~~ Accounting (FCTR)
- x Final Report of Inventions
- _ Govt. Property Inventory & Related Certificate
- _ Classified Material Certificate
- _ Other ______

Assigned to: College of Management (School/~~Laboratory~~)

COPIES TO:

Project Director
Division Chief (EES)
School/Laboratory Director
Dean/Director—EES
Accounting Office
Procurement Office
Security Coordinator (OCA)
~~Reports~~ Coordinator (OCA)
Library, Technical Reports Section
EES Information Office
Project File (OCA)
Project Code (GTRI)
Other ______

CA-4 (1/79)

Reprinted from JOURNAL OF ECONOMIC THEORY
Vol. 23, No. 2, October 1980

Notes, Comments, and Letters to the Editor

Capital Accumulation and the Optimization of Renewable Resource Models*

ROY MENDELSSOHN

Southwest Fisheries Center, National Marine Fisheries Service, NOAA, Honolulu, Hawaii 96812

AND

MATTHEW J. SOBEL

College of Industrial Management, Georgia Institute of Technology, Atlanta, Georgia 30332

Received January 23, 1978; revised August 27, 1979

1. INTRODUCTION

There are separate literatures on normative models of capital accumulation, fisheries management, and reservoir operation. However, generic models of each type share a common mathematical structure. The models often include uncertainty in the length of time to which planning should apply and in the future consequences of present decisions.

The present paper investigates the structure of optimal decision making under uncertainty for general single sector growth models, for individual optimal consumption and savings models, for models of a single fish species with pooled age classes, and for models of a single reservoir. The results extend and unify some of those of Amir (1967), Bewley (1977), Levhari and Srinvinasan (1969), Hakaanson (1970), Mirman (1971), Brock and Mirman (1972, 1973), Miller (1974), Sobel (1975), Whitt (1975a), Schechtman (1976), Schechtman and Escudero (1977), and Yaari (1976). Also we present new and simpler proofs for most of the theorems which generalize results of the authors above.

For brevity, we use the terminology of capital accumulation. Let the first consumption and reinvestment decisions be made in period n and the last

* Partially supported by U.S. National Science Foundation Grants GK-13757, EnG 76-15559, and SOC 78-05770. Parts of this paper appeared in Dr. Mendelssohn's Ph. D. thesis, Yale University, 1976.

0022-0531/80/050243-18$02.00/0

ones in period 1. At the beginning of each period t ($t = 1, 2,\ldots$) let x_t denote the capital on hand in units of dollars or physical quantities as the context dictates. The decisions in period t are y_t, namely the amount of x_t that is reinvested, and $z_t = x_t - y_t$, the amount consumed. Let

$$y_t \in Y(x_t) \qquad (t = 1, 2,\ldots) \tag{1.1}$$

constrain the reinvestment and consumption decisions. The connection between reinvestment decisions and accumulated capital is

$$x_{t+1} = s(y_t, D_t), \tag{1.2}$$

where $D_1, D_2,\ldots, D_n$ are assumed to be independent random variables that are distributed as the generic random variable D. We assume for each t that x_t lies in a convex reference set X and that $C \equiv \{(x, y): y \in Y(x), x \in X\}$ is a convex set.

Let $G(x_t, y_t)$ denote the utility in period t of having an initial capital x_t and reinvesting y_t. Let α denote the single-period discount factor. The generic problem is maximization of expected discounted utility, namely $E \sum_{t=1}^{n} \alpha^{t-1} G(x_t, y_t)$, where $n \leqslant \infty$. With a *consumption horizon* of n periods until termination, let $A_n(x)$ denote an optimal reinvestment decision and $x - A_n(x)$ an optimal consumption decision. (For a fishery model, $x - A_n(x)$ is the amount harvested, and $A_n(x)$ is the population size after harvesting ceases. For a reservoir model, x is the amount of water in the reservoir, and $x - A_n(x)$ is the amount discharged.)

The real line is indicated by $\mathbf{R}$, and $\mathbf{R}_+$ denotes $[0, \infty)$. If $z \in \mathbf{R}$ then $(z)^+$ denotes $\max(0, z)$. Derivatives (partial or regular) are from the left when necessary. If $w(\cdot, \cdot)$ is a function of two variables, then $w^{(1)}(u, v)$ and $w^{(2)}(u, v)$ denote the partial derivatives with respect to the first and second arguments. If $a \in \mathbf{R}$ and $b \in \mathbf{R}$, then $a \wedge b$ denotes $\min(a, b)$.

Suppose w is a real-valued concave function on $\mathbf{R}$ and D is a random variable for which the expectations $r(z) \equiv Ew(z - D)$ and $Ew'(z - D)$ both exist. Then $r'(z) = Ew'(z - D)$ can be justified by the monotonicity in δ of $|w(z - D) - w(z - D - \delta)|/\delta$ (because w is concave) and the Dominated Convergence Theorem (Royden 1963).

2. New Features of the Model

The results in this paper relax the assumptions of the previously cited papers in five important ways. First, many of our results do not assume that single-period utility depends only on consumption, that is $G(x, y) \equiv g(x - y)$. In renewable resource models, the benefits of a harvest $x_t - y_t$ are offset by harvesting costs that typically depend separately on the stock size x_t and the amount harvested $x_t - y_t$.

Second, when the utility function does depend only on consumption, we assume neither $g'(0) = +\infty$ nor strict concavity of $g(\cdot)$. Neither assumption is salient for many renewable resource models. However, one or both of the assumptions is found in most of the cited references, most particularly Bewley (1977), Brock and Mirman (1972), Mirman and Zilcha (1975), and Schechtman (1976). Thus their models fail to encompass either linear utilities or quadratic utilities, or both. These cases are empirically useful.

Third, we avoid assumptions about $s(\cdot, \cdot)$ (in (1.2)) that would block applications to renewable resource models. The Inada condition

$$s^{[1]}(0, \cdot) \equiv +\infty, \qquad s^{[1]}(\infty, \cdot) \equiv 0$$

are not imposed. Note that the simple case $s(y, d) = \rho y + d$ violates the Inada condition as do quadratic functions. For several of the results it is not assumed that $s(\cdot, d)$ is nondecreasing for each fixed value d of D_t. We avoid continuity and ordering assumptions about $s(y, \cdot)$ such as are found in Brock and Mirman (1972), Schechtman (1976), and Bewley (1977). When $G(x, y) = p \cdot (x - y)$, new results are presented that assume only pseudoconcavity of $s(\cdot, d)$ for each fixed value d of D_t. This last assumption is often made in models of renewable resources.

Fourth, *effective constraints* are allowed in the model. We do not assume existence of interior solutions. In renewable resource models, the absence of Inada conditions typically causes some constraints to be active at an optimum. This makes the analysis more complex because it is no longar possible to assume that the derivative of an optimal value function equals $g'(x, x - A_n(x))$. The methods of proof, therefore, differ significantly from those in previously cited papers.

Finally, a unified treatment is presented for both the finite and infinite horizon problems. Our proof that the infinite horizon optimal policy $A(x)$ is the limit of the $A_n(x)$'s is more straightforward than those of Schechtman and Escudero (1977) (who assume $s(y_t, D_t) = \rho y_t + D_t$) and Brock and Mirman (1972) (who assume strict concavity, interior solutions and impose other restrictive conditions) although our model is sufficiently general to encompass renewable resource applications.

Also, we present a straightforward short proof that a unique stationary distribution of wealth exists. The proof uses a result of Rosenblatt (1967) and is much shorter and simpler than proofs of essentially the same theorem by Brock and Mirman (1972), Brock and Majumdar (1975), and Mirman and Zilcha (1975).

The model in Bewley (1977) is richer probabilistically than ours. Although his model requires $G(x, y) \equiv g(x - y)$ and $s(y, d) \equiv y + d$, g is stochastic with $\{(g_t, D_t)\}$ being a stationary stochastic process.

3. Assumptions

The following assumptions are made in various combinations.

(3.1) G is finite, concave, and continuous on C;

(3.2) for each y, $G(\cdot, y)$ is nondecreasing on $\{x: (x, y) \in C\}$;

(3.3) $G(x, y+\gamma) - G(x, y) \leqslant G(x+\lambda, y+\gamma) - G(x+\lambda, y)$ for all $\gamma > 0$, $\lambda > 0$ with all arguments in C;

(3.4) G is nonnegative and continuous on C;

(3.5) for each d, $s(\cdot, d)$ is continuous and concave on the set $\mathbf{Y} \equiv \bigcup_{x \in X} Y(x)$;

(3.6) for each d, $s(\cdot, d)$ is continuous and concave on the set $[0, y^0(d)]$ and convex on the set $[y^0(d), \infty)$;

(3.7) for each d, $s(\cdot, d)$ is nondecreasing on $\mathbf{Y}$;

(3.8) $G(x, y) = p \cdot (x - y), p \geqslant 0$;

(3.9) $Y(x) = [0, x], x \in X$.

Assumption (3.4) is slightly deceptive because several articles, including Phelps (1962), Hakaason (1970), and Miller (1974), have $G(x, y) = \log(x - y)$, for $x - y \geqslant \xi > 0$. However, nonnegativity of G is equivalent to having a uniform lower bound. Suppose, for example, that $G(x, y) \geqslant -B$ if $(x, y) \in B$ $(B > 0)$, and let $G^*(x, y) = G(x, y) + B$. Then a policy is optimal with utility function G^* if and only if it is optimal also for G. Moreover, G^* is nonnegative on C. Assumption (3.3) is equivalent to *supermodularity* of G on C. Topkis (1978) discusses supermodularity and its consequences for optimization.

4. Optimal Policies

From standard dynamic programming arguments, the generic problem of maximizing $E \sum_{t=1}^{\infty} \alpha^{t-1} G(x_t, y_t)$ leads to the recursion

$$f_n(x) = \sup\{J_n(x, y): y \in Y(x)\}, \qquad x \in X, \tag{4.1}$$

$$J_n(x, y) = G(x, y) + \alpha E f_{n-1}(s[y, D]), \qquad y \in Y(x), \quad x \in X \tag{4.2}$$

for $n = 1, 2, \ldots$ with $f_0(\cdot) \equiv 0$. The proofs of most results will exploit the following diminishing returns property of $f_n(\cdot)$.

THEOREM 4.1. *Assumptions* (3.1), (3.2), (3.5), *and* (3.9) *imply for each n that $f_n(\cdot)$ is continuous, concave, and nondecreasing on X.*

Proof. The proof that $f_n(\cdot)$ inherits concavity and is nondecreasing is well known and will not be reproved. Our proof of continuity is new and simpler than related proofs in Brock and Mirman (1972) and Schechtman (1976).

Let p be a metric on the space in which X lies. To initiate a contrapositive proof, suppose f_n experiences a discontinuity at $x^0 \in X$. Then there is a $y^0 \in Y(x^0)$ and a real number $\gamma > 0$ with the property that for all real numbers $\delta > 0$ there exists $x \in X$ such that

$$p(x, x^0) < \delta \qquad \text{and} \qquad J_n(x, y) \geqslant \gamma + J_n(x^o, y^o), \qquad y \in Y(x). \quad (4.3)$$

(Concavity precludes the reverse inequality for J). Because C is a convex set it is possible to select a subsequence (x^j, y^j) satisfying (4.3) and $(x^j, y^j) \to (x^0, y^0)$. Therefore, contrary to assumption, J_n is discontinuous at (x^0, y^0). ∎

Continuity of $J_n(x, \cdot)$ on $Y(x)$ for each $x \in X$ follows from (3.1), (3.5), and continuity of $f_{n-1}(\cdot)$ via Theorem 4.1. Compactness of $Y(x)$ via (3.9), therefore, implies attainment of the supremum in (4.1) for each $x \in X$ and existence of an optimal policy.

The next two theorems describe the dependence of $A_n(x)$ on x. They lead (in Corollary 4.2) to sufficient conditions for $0 \leqslant dA_n(x)/dx \leqslant 1$. Many authors have proved at least one side of this inequality for special cases of our model.

THEOREM 4.2. *Assumptions* (3.1)–(3.3), (3.5), *and* (3.9) *imply for each* n *that there exists* $A_n(\cdot)$ *with the property*

$$A_n(x') \geqslant A_n(x) \qquad \text{if} \quad x' \geqslant x. \quad (4.4)$$

Proof. The theorem would be true if, for $\delta > 0$,

$$J_n(x', A_n(x)) - J_n(x', A_n(x) - \delta) \geqslant J_n(x, A_n(x)) - J_n(x, A_n(x) - \delta) \geqslant 0.$$

The right inequality is implied by the optimality of $A_n(x)$. The left inequality would be implied by $J_n^{[2]}(\cdot, y)$ being nondecreasing on $\{x: (x, y) \in C\}$. From (4.2),

$$J_n^{[2]}(x, y) = G^{[2]}(x, y) + \alpha E\{f'_{n-1}(s[y, D])\, s^{[1]}(y, D)\}$$

so (3.3) completes the proof. ∎

Theorem 4.2 requires neither $G(x, y) \equiv g(x - y)$ nor strict concavity. It assumes concavity of $s(\cdot, d)$ but not monotonicity. Theorem 4.3 obtains a further result when $G(x, y) \equiv g(x - y)$ and $Y(x) = [0, x]$.

THEOREM 4.3. *Assumptions* (3.1)–(3.3), (3.5), (3.9), *and* $G(x, y) \equiv g(x-y)$ *imply for each* n *that there exists* $A_n(\cdot)$ *with the property*

$$0 \leqslant A_n(x') - A_n(x) \leqslant x' - x \qquad \text{if} \quad x \leqslant x' \tag{4.5}$$

Proof. The left inequality is (4.4). To prove the right one, let $z = x - y$ and rewrite (4.1) and (4.2) as

$$f_n(x) = \sup\{H_n(x, z): 0 \leqslant z \leqslant x\}, \qquad x \in X,$$
$$H_n(x, z) = g(z) + \alpha E f_{n-1}(s[x-z, D]).$$

The right side of (4.5) would be valid if

$$\begin{aligned} 0 &\leqslant H_n^{[2]}(x+\delta, z) - H_n^{[2]}(x, z) \\ &= \alpha E\{f'_{n-1}(s[x-z, D])\, s^{[1]}(x-z, D) \\ &\quad - f'_{n-1}(s[x+\delta-z, D])\, s^{[1]}(x+\delta-z, D)\} \end{aligned} \tag{4.6}$$

for $\delta > 0$ as will be shown.

Concavity of $s(\cdot, d)$ implies

$$s^{[1]}(x+\delta-z, d) \leqslant s^{[1]}(x-z, d).$$

If $s^{[1]}(x+\delta-z, d) \geqslant 0$ then concavity implies $s(x-z, d) \leqslant s(x+\delta-z, d)$ so $0 \geqslant f'(s[x+\delta-z, d]) \leqslant f'(s[x-z, d])$ and

$$f_{n-1}(s[x+\delta-z, d])\, s^{[1]}(x+\delta-z, d) \leqslant f_{n-1}(s[x-z, d])\, s^{[1]}(x-z, d). \tag{4.7}$$

If $s^{[1]}(x+\delta-z, d) < 0 \leqslant s^{[1]}(x-z, d)$ then (4.7) is trivial. If $0 > s^{[1]}(x-z, d) \geqslant s^{[1]}(x-z+\delta, d)$ then $s(x-z, d) \geqslant s(x+\delta-z, d)$ because $s(\cdot, d)$ is concave. Concavity and monotonicity of $f_{n-1}(\cdot)$ and these inequalities yield

$$0 \leqslant f'_{n-1}(s[x-z, d]) \leqslant f'_{n-1}(s[x+\delta-z, d]),$$
$$0 > s^{[1]}(x-z, d) \geqslant s^{[1]}(x+\delta-z, d)$$

and, therefore, (4.7), which proves (4.6). ∎

COROLLARY 4.1. *The assumptions of Theorem* 4.3 *and* $0 < A_n(x') < x'$ *for some* $x' > 0$ *imply* $0 < A_n(x)$ *for all* $x \geqslant x'$.

COROLLARY 4.2. *The assumptions of Theorem* 4.3 *imply*

$$0 \leqslant dA_n(x)/dx \leqslant 1, \qquad x \in X^0, \tag{4.8}$$

where X^0 *denotes the interior of* X.

Proof. From Theorem 4.4, $A_n(\cdot)$ is nondecreasing so its discontinuities, if any, are upward jumps. These jumps are precluded by $A_n(x)-x$ being nonincreasing so $A_n(\cdot)$ is continuous. Monotonicity of $A_n(\cdot)$ implies differentiability except, possibly, on a set of measure zero where one-sided derivatives exist, so (4.6) implies (4.8). ∎

Corollary 4.2 treats a more general problem than Schechtman (1976) does and its proof seems more straightforward.

An optimal policy, $A_n(x)$ can be described in further detail if $Y(x)=[0,x]$ and $G(x,y)=p\cdot(x-y)$ for $p\geqslant 0$. After substitution and rearrangement of terms, the optimization problem becomes:

$$\text{maximize } E\left\{px_1-\alpha^{T-1}y_T+p\sum_{t=1}^{T}\alpha^{t-1}(\alpha s[y_t,D_t]-y_t)\right\}$$

subject to $0\leqslant y_t\leqslant x_t$, $t=1,\dots,T$. The first term, px_1, is fixed. The second term, $-\alpha^{t-1}y_T$, has $y_T\equiv 0$ for an optimal policy (if $\alpha>0$). Therefore, an equivalent problem, in the sense of having the same optimal policy for all $n>1$, is given by the following recursion:

$$f_0(\cdot)\equiv 0,$$

$$f_n(x)=\sup\{J_n(y)\colon 0\leqslant y\leqslant x\},$$

where $J_n(y)=G(y)+\alpha Ef_{n-1}(s[y,D])$ and $G(y)=p\cdot(\alpha Es[y,D]-y)$.

Let x_n^0 denote a global maximum of $J_n(y)$. For $x_n\geqslant x_n^0$, it is optimal to consume $x_n-x_n^0$. If $J_n(\cdot)$ is pseudoconcave for all n, then it is straightforward to prove that an optimal policy is given by:

$$A_n(x)=x\wedge x_n^0.$$

What conditions ensure pseudoconcavity of $J_n(y)$? Corollary 4.3 is an immediate extension of Theorem 4.1.

COROLLARY 4.3. *Assumptions* (3.4), (3.5), (3.8), *and* (3.9), *imply*:

(i) $J_n(y)$ *is concave and continuous,*

(ii) $A_n(x)=x\wedge x_n^0$. ∎

Suppose

$$s(y,d)=d\phi(y),\qquad P\{D\geqslant 0\}=1,\tag{4.9}$$

and

$\phi(\cdot)$ *is pseudoconcave and continuous with mode at* y_m.

When both (3.6) and (4.9) are valid, $y^0(\cdot)\equiv y^0\geqslant y_m$.

THEOREM 4.4. *Assumptions* (3.6), (3.9) *and* (4.9) *imply*

$$A_n(x) = x \wedge x_n^0$$

is optimal.

Proof. Let $\mu = E(D)$. At $n = 1$, $x_1^0 = \inf\{y: \phi'(y) \leqslant (\alpha\mu)^{-1}\} \leqslant y^m \leqslant y^0$. We use $f_0(\cdot) \equiv 0$. The inductive assumption is that $f_{n-1}(\cdot)$ is concave nondecreasing on X. In

$$J_n'(y) = p(\alpha\mu\phi'[y] - 1) + \alpha\phi'(y)\,E(Df_{n-1}'[D\phi(y)]),$$

the first term is nonpositive if $y \geqslant y^m \geqslant x_1^0$ and the second term is nonpositive if $y \geqslant y^m$ because then $\phi'(y) \leqslant 0$ (while $f'(\cdot) \geqslant 0$ and $P\{D \geqslant 0\} = 1$). Therefore, $x_n^0 \leqslant y^m$ so $A_n(x) = x \wedge x_n^0$ and $f_n(x) = J_n(x \wedge x_n^0)$. It follows that $f_n(\cdot)$ is concave nondecreasing if $J_n(\cdot)$ is concave on $[0, x_n^0]$, which is now verified.

$$J_n(y) = p(\alpha\mu\phi[y] - y) + \alpha E(f_{n-1}[D\phi(y)]),$$

whose first term is concave on $[0, x_n^0]$ because $\phi(\cdot)$ is concave on $[0, y^0]$ and $x_n^0 \leqslant y^m \leqslant y^0$. Concavity of the second term is implied by the inductive assumption, $\phi(y)$ being concave nondecreasing on $[0, x_n^0]$, and $P\{D \geqslant 0\} = 1$. ∎

Theorem 4.4 shows that, if $s(y, d) = d\phi(y)$, the shape of $\phi(\cdot)$ beyond its mode doesn't effect an optimal policy in any significant way because an optimal policy always returns the state to the concave part of the curve.

The properties assumed for $s(\cdot, \cdot)$ can be relaxed by making further assumptions about the distribution of D. A stochastic kernel $K(x, y)$ is TP_2 if, for all $x_1 < x_2$ and $y_1 < y_2$, $K(x_1, y_1)\,K(x_2, y_2) \geqslant K(x_2, y_1)\,K(x_1, y_2)$. TP_2 kernels include the exponential and range families of densities which contain the binomial, Poisson, gamma, and normal (with fixed variance) densities.

THEOREM 4.5. *Assumptions* (3.9), (3.8) *with* $p \geqslant 0$, $s(\cdot, d)$ *pseudoconcave and continuous for each d, and D with a continuous density function that is* TP_2 *implies*:

$$A_n(x) = x \wedge x_n^0.$$

Proof. This theorem and Theorem 4.1 have similar proofs except it must be shown (a) that a convex combination (expectation) of pseudoconcave functions using a random variable with a TP_2 density is again pseudoconcave, and (b) that a nondecreasing nonnegative pseudoconcave function of a pseudoconcave function is again pseudoconcave.

The first claim is Theorem 5.1 of Chapter 3 in Karlin (1968). The second claim is proven here for differentiable functions. From the definition of pseudoconcave functions, $f(\phi[\cdot])$ is pseudoconcave if

$$f'(\phi[y])\,\phi'(y)(y'-y) \leqslant 0 \qquad \text{implies} \quad f(y') \leqslant f(y)$$

First, $f'(\cdot) \geqslant 0$ so the only pertinent case is $\phi'(y)(y'-y) \leqslant 0$. The pseudoconcavity of $\phi(\cdot)$ implies $\phi(y') \leqslant \phi(y)$. However, $f(\cdot)$ is nondecreasing and pseudoconcave, so $f(\phi[y']) \leqslant f(\phi[y])$. ∎

5. Effects of the Consumption Horizon

This section investigates the impact of the consumption horizon on the structure of an optimal policy and on its valuation. The following result presents sufficient conditions for a longer consumption horizon to induce a higher valuation, greater accumulation, and higher incremental benefits per unit of added capital.

THEOREM 5.5. *For each n and $x \in X$:*

(a) *Assumption* (3.4) *implies*

$$f_n(x) \leqslant f_{n+1}(x); \tag{5.1}$$

(b) *Assumptions*[1] (3.2), (3.3), (3.4), (3.7) *and* (3.9) *imply*

$$A_n(x) \leqslant A_{n+1}(x); \tag{5.2}$$

$$J_n(x, y+\gamma) - J_n(x, y) \leqslant J_{n+1}(x, y+\gamma) - J_{n+1}(x, y), \quad \gamma > 0 \quad \textit{if} \quad (x, y) \in C \quad \textit{and} \quad (x, y+\gamma) \in C; \tag{5.3}$$

$$f_n(x+\gamma) - f_n(x) \leqslant f_{n+1}(x+\gamma) - f_{n+1}(x) \qquad \textit{if} \quad x+\gamma \in X, \quad \gamma > 0. \tag{5.4}$$

Proof. (a) $f_0(\cdot) \equiv 0$ initiates a straightforward inductive proof of (5.1) that uses (3.4).

(b) Observe that (5.3) is supermodularity (cf. Topkins (1978)) of $J_n(x, y)$ in (y, n) for each x, and (5.4) is supermodularity of $f_n(x)$ in (x, n). If $r(a, b)$ is supermodular in (a, b) and $M(\cdot)$ is nondecreasing then $r(a, m[b])$ also is supermodular in (a, b). Hence, *if* $f_{k-1}(x)$ is supermodular in (x, k) for all $k \leqslant n-1$ then (4.2) and (3.7) imply supermodularity of $J_k(x, y)$ in (y, k) for all $k \leqslant n$.

If $J_k(x, y)$ is supermodular in (x, y, k) for all $k \leqslant n$ and if C is a lattice

[1] Instead of (3.9) and convexity of X, it is sufficient to assume that $Y(x)$ is a compact lattice for each x, C and X are lattices, and $Y(x)$ is *ascending* on X.

then Theorem 6.2 in Topkis (1978) implies $A_{k-1}(x) \leqslant A_k(x)$ for all $k \leqslant n$. Assumption (3.9) and convexity of X imply that C is a lattice so it remains to establish (5.4).

Assumption (3.2) implies $f_1(x+y) - f_1(x) \geqslant 0$ (Theorem 1) so $f_0(\cdot) \equiv 0$ implies (5.4) is valid for $n = 0$. Inductively, if (5.4) is valid for all $n \leqslant k-1$ then (5.3) is valid for all $n \leqslant k$. Then Theorem 4.3 in Topkis (1978) implies (5.4) for all $n \leqslant k$. ∎

We are grateful to Donald M. Topkis of Bell Laboratories for suggesting this line of proof. Our earlier version of Theorem 5.1(b) contained superfluous assumptions including concavity of G.

The next result concerns the limiting behavior of f_n and A_n as $n \to \infty$. Suppose

$$0 < \alpha < 1; \tag{5.7}$$

$$G^{[1]}(x, y) < \infty, \qquad (x, y) \in C; \tag{5.8}$$

$$G(x, \cdot) \text{ is nonincreasing on } Y(x), \qquad x \in X; \tag{5.9}$$

let $r_1(x) \equiv x$ and $r_{n+1}(x) \equiv s(r_n(x), D_n)$; then $E \sum_{n=1}^{\infty} \alpha^{n-1} r_n(x) < \infty$. (5.10)

THEOREM 5.2. (a) *Assumptions* (3.1)–(3.5), (3.7), *and* (3.9) *imply for each* $x \in X$ *existence of*

$$A(x) = \lim_{n \to \infty} A_n(x). \tag{5.11}$$

If $G(x, y) \equiv g(x-y)$ *then*

$$0 \leqslant A(x') - A(x) \leqslant x' - x, \qquad x \leqslant x. \tag{5.12}$$

(b) *Assumptions* (3.1), (3.2), (3.4), (3.5), (3.7), (3.9), *and* (5.7)–(5.10) *imply for each* $x \in X$ *existence of*

$$f(x) \equiv \lim_{n \to \infty} f_n(x) \tag{5.13}$$

with $f(\cdot)$ *being concave and nondecreasing on* X.

Proof. (a) From (5.2) $A_n(x) \leqslant A_{n+1}(x) \leqslant x$ for every n and x. Hence, monotone convergence yields (5.11) and, via Theorem 4.3, it yields (5.12).

(b) Optimality of $A_n(x)$, (5.9), (3.7), and $f_{n-1}(\cdot)$ nondecreasing imply

$$\begin{aligned} f_n(x) &= G(x, A_n(x)) + \alpha E f_{n-1}(s[A_n(x), D_n]) \\ &\leqslant G(x, 0) + \alpha E f_{n-1}(s(x, D_n)) \\ &\leqslant \sum_{t=1}^{n} \alpha^{t-1} E G(r_t(x), 0). \end{aligned}$$

Concavity of G implies

$$G(u,v) \leqslant G(u^0,v^0) + [G^{[1]}(u^0,v^0), G^{[2]}(u^0,v^0)] \begin{pmatrix} u-u^0 \\ v-v^0 \end{pmatrix}$$

for all (u^0, v^0) and $(u,v) \in C$. However, $G^{(2)} \leqslant 0$ from (5.9) so

$$\begin{aligned} f_n(x) &\leqslant \sum_{t=1}^{n} \alpha^{t-1} E[G(x,0) + G^{[1]}(x,0)(r_t(x) - x)] \\ &\leqslant [G(x,0) - xG^{[1]}(x,0)]/[1-\alpha] + G^{[1]}(x,0) \sum_{t=1}^{\infty} \alpha^{t-1} r_t(x), \end{aligned}$$

which is finite from (5.8) and (5.10). Therefore, $f_1(x), f_2(x), \ldots,$ is a bounded monotone sequence which implies (5.13). Monotone convergence implies that $f_1(\cdot), f_2(\cdot), \ldots,$ endow $f(\cdot)$ with their properties of concavity and monotonicity. ∎

The next result uses a familiar argument from inventory theory (Sobel 1970a) to prove that $f(\cdot)$ satisfies a functional equation analogous to (4.1). Let

$$J(x,y) = G(x,y) + \alpha E f(s[y,D]),$$

which exists by virtue of the following proof. The proof is much simpler than those of similar theorems in Brock and Mirman (1972) and Schechtman (1976).

THEOREM 5.3. *Assumptions* (3.1), (3.2), (3.4), (3.5), (3.7), (3.9), *and* (5.7)–(5.10) *imply*

$$\begin{aligned} f(x) &= \sup\{J(x,y) : y \in Y(x)\}, \qquad x \in X, \\ &= J(x, A[x]). \end{aligned} \tag{5.14}$$

Proof. For all x and n, $f_{n-1}(x) \leqslant f_n(x)$ so

$$J_{n-1}(x,y) \leqslant J_n(x,y) \leqslant f(x) \leqslant B(x) < \infty, \qquad y \in Y(x), \quad x \in X,$$

where $B(x)$ is the bound developed in the proof of (b) in Theorem 5.2. Therefore, $\{J_n(x,y)\}$ is a bounded monotone sequence so $J_n(x,y) \leqslant J(x,y)$ and

$$f_n(x) = \sup\{J_n(x,y) : y \in Y(x)\} \leqslant \sup\{J(x,y) : y \in Y(x)\}.$$

Convergence of f_n to f implies

$$f(x) \leqslant \sup\{J(x,y) : y \in Y(x)\}, \qquad x \in X,$$

so the theorem will have been proved after establishing

$$f(x) \geqslant \sup\{J(x, y): y \in Y(x)\}, \qquad x \in X. \tag{5.15}$$

Monotone convergence implies

$$f(x) \geqslant f_n(x) = \sup\{J_n(x, y): y \in Y(x)\},$$
$$f(x) \geqslant \lim_{n \to \infty} \sup\{J_n(x, y): y \in Y(x)\},$$

whereas the right side of (5.15) is

$$\sup\{\lim_{n \to \infty} J_n(x, y): y \in Y(x)\}.$$

Therefore, for (5.15) it is sufficient to prove

$$\lim_{n \to \infty} \sup\{J_n(x, y): y \in Y(x)\} = \sup\{\lim_{n \to \infty} J_n(x, y): y \in Y(x)\}. \tag{5.16}$$

The existence of the limit on the left side of (5.16) is implied by (5.13). For each n, $J_n(x, \cdot)$ is continuous on $Y(x) = [0, x]$ so $J_n(x, \cdot) \to J(x, \cdot)$ uniformly on $[0, x]$ because the Dominated Convergence Theorem implies $Ef_{n-1}(s[\cdot, D]) \to Ef(s[\cdot, D])$. Therefore,

$$0 = \lim_{n \to \infty} \sup\{J(x, y) - J_n(x, y): 0 \leqslant y \leqslant x\},$$

which implies (5.16) and, consequently, (5.15). ∎

$A(\cdot)$ inherits the properties of $\{A_n(\cdot)\}$.

COROLLARY 5.1. *The assumptions of Theorem* 5.2(a) *imply*

$$0 \leqslant A'(x) \leqslant 1. \tag{5.17}$$

6. ACCUMULATION USING A STATIONARY POLICY

Suppose the same policy $A(\cdot)$, arbitrary and possibly suboptimal, but satisfying (5.12), is used each period. Then reinvestment and consumption each period t are given by $A(\chi_t)$ and $\chi_t - A(\chi_t)$, where χ_t denotes the random asset level at the beginning of period t. Successive asset levels are connected by

$$\chi_{t+1} = s[A(\chi_t), D_t], \qquad t = 1, 2, \ldots, \tag{6.1}$$

or, equivalently, by a kernel $K(A(x), \Gamma)$ which is the probability of being in

$\Gamma \subseteq X$ if $A(x)$ is the action taken at state $x \in X$ (cf. Feller (1971)). More formally, consider the state space X with a Borel field of subsets B. Then $K(\cdot, \cdot)$ is a probability kernel if $K(x, \cdot)$ is a probability measure on B for each $x \in X$ and $K(\cdot, \Gamma)$ is a B-measurable function for each $\Gamma \in \beta$.

Whether or not the Markov processes χ_t converge to a stationary distribution is an important question. Convergence results in certain cases are given in Brock and Mirman (1972), Brock and Majumdar (1975), Mirman and Zilcha (1975), Schechtman (1976), and Schechtman and Escudero (1977). By comparison with these papers, the approach here relies on properties of the stochastic kernel $K(\cdot, \cdot)$. Therefore, the proofs are more direct and do not in any essential way depend on scalar properties of x_t and y_t (although the theorems are presented only for this case).

$K(\cdot, \cdot)$ induces the following operator T which takes bounded measurable functions h into bounded measurable functions:

$$(Th)(x) = \int K(x, dy)\, h(y). \tag{6.2}$$

There is a dual representation which takes probability measures Q into probability measures; namely, for each $\Gamma \subseteq X$,

$$(VT)(\Gamma) = \int Q(dx)\, K(x, \Gamma) \tag{6.3}$$

(Rosenblatt 1967). The operator T is equicontinuous if it maps continuous functions into continuous function. For the remainder of this section, suppose X is a compact set, $0 \in X$, and $s(\cdot, d)$ is continuous; note that (5.12) implies continuity of $A(\cdot)$. Therefore, it is straightforward to show that $K(\cdot, \cdot)$ induces an equicontinuous family of transformations. Consider the following three conditions:

(A) $K(x, I) > 0$ for all open intervals $I \subseteq X$.

(B) There exists a compact subset L of X, such that for each $x \in L$, $K(x, L) = 1$ and the operator T defined in (6.3) is irreducible on L in the sense of Rosenblatt (1967, p. 476).

(C) Neither (A) nor (B) holds.

Condition (A) is satisfied, for example, by models where $s(\cdot, \cdot)$ is linear, as in Schechtman and Escudero (1977), where $s(y, d) = ry + d$. Condition (B) is often proven as a preliminary result, as in Section 3 of Brock and Mirman (1972). Condition (C) is important when determining if a stationary distribution concentrates all its mass at a single point. The three conditions separate the question of the existence of a unique stationary distribution from the question of the existence of at least one, perhaps many, invariant measures on the set X, given some policy $A(x)$. Some of the literature confuses the two questions.

THEOREM 6.1. *If X is compact and $K(\cdot, \cdot)$ induces an equicontinuous family of transformatrions, then*:

(i) *There is at least one invariant measure on X.*

(ii) (A) *implies there is only one invariant measure on X and it is the unique positive stationary distribution on X.*

(iii) (B) *implies there is a unique invariant measure for each closed irreducible subset L of X, and this measure is the unique positive stationary distribution on L.*

(iv) *Suppose also, for some $\gamma > 0$, $P\{s(A[x], D) \geqslant x\} = 1$ for all $x \in [0, \gamma]$. Let $x^* = \sup\{x: x \in X\}$ and suppose $L = \{x: \gamma \leqslant x \leqslant x^*\}$ is irreducible. Then there is a unique stationary measure on X which has positive probability only on open intervals that intersect with L.*

(v) *If 0 is an absorbing state then* (C) *implies that there is a unique stationary distribution concentrated at 0 (so $P\{\liminf x_t = 0\} = 1$ and the process tends to gets arbitrarily close to 0).*

Proof. Part (ii) is Theorem 2 in Feller (1971, p. 272). Parts (i), (iii), (iv), and (v) are implied by Theorems 3 and 4 in Rosenblatt (1967) and Theorem 2.4 in Jamison (1964). ∎

Boylan (1977) proves theorems similar to (iv) and (v) although his approach is different. Two examples illustrate the power of the theorem. First, for the models in Schechtman (1976) and Schechtman and Escudero (1977), part (ii) of the theorem immediately implies convergence to a unique stationary distribution. Second, the model in Brock and Mirman (1972) has an equicontinuous kernel on a compact set so it necessarily possesses an invariant measure. This avoids the lengthy argument in Section 4 of Brock and Mirman's paper. Moreover, our results depend on irreducibility of an operator on a *subset* of the set of states so one can focus on states that "communicate" rather than on fixed points of growth functions. This permits simplified proofs of the results in Section 3 of Brock and Mirman's paper. Boylan (1977) gives such a simplified proof.

7. EXTENSIONS

The results thus far concern a model whose structure is stationary over time and in which growth does not occur. These restrictions can be relaxed.

Suppose, for example, that utility alters with age or time as well as with wealth. Let $G_k(x, y)$ denote the utility function appropriate to consumption at age $n-k$. Then all results in Section 4 and Theorem 5.1 remain valid if G is replaced by G_n in the assumptions in Section 3. As one example that generalizes Hakaanson (1973), let $G_k(\cdot, \cdot) \equiv 0$ if $k > 1$.

If the random variables $D_n, D_{n-1},..., D_2, D_1$ exhibit a dependence, let $T_k(D_n,..., D_{n-k+1})$ denote a statistic of $D_n,..., D_{n-k+1}$ that is sufficient for D_t. Let ϕ_k be a function which maps (T_k, D_{n-k}) into T_{k-1}. Such functions must exist, because the entire past history is a sufficient statistic. Then the results in Sections 4 and 5 through Theorem 5.1 remained unchanged if each $Y_n(x)$ is a convex set and $Y_n(x) \subseteq Y_n(x')$ if $x' \geqslant x$ for each n. If $Y_n(x)$ is appropriately convergent in n for each x, then generalization of Theorems 3.2, 5.3, and Corollary 5.1 can be obtained.

Exogenous price processes arise when there is a random sequence of prices $\rho_1, \rho_2, \rho_3,...,$ such that

$$G(x, y, \rho) = u(\rho[x-y])$$

and the constraint set is now $Y(x, p)$ for each fixed value p of ρ. If the assumptions in Section 3 are valid for each fixed value p of ρ, then the results in Sections 4 and 5 are true for each fixed p. For example, Theorem 5.1 becomes $f_{n-1}(x, p) \leqslant f_n(x, p)$, $A_{n-1}(x, p) \leqslant A_n(x, p)$, etc.

Undiscounted Utilities

The results in Section 4, Theorems 5.1 and 6.1 are true for all values of the discount factor $\alpha \geqslant 0$. If $\alpha \geqslant 1$ then generally $f_n(x) \to \infty$ as $n \to \infty$ so Theorems 5.2b and 5.3 are uninteresting. The validity of Theorem 5.2a is an intriguing issue when $\alpha = 1$.

The overtaking criterion (see Brock and Mirman (1973); Brock and Majumdar (1975) and their references to the work of Von Weizsacker, Gale, and Gale's students; Denardo and Rothblum (1979)) is currently in vogue when utilities are undiscounted. Consider, instead, the average gain criterion from the theory of Markov decision processes, namely,

$$\Gamma(Z \mid x) \equiv \varlimsup_{T \to \infty} \sum_{j=1}^{T} G(x_j, y_j)/T \tag{7.2}$$

and search for a *gain-optimal* policy Z^* such that

$$\Gamma(Z^* \mid x) = \sup_Z \Gamma(Z \mid x) \qquad \text{for all} \quad x \in X.$$

It can be shown that a policy is gain optimal if it is overtaking optimal (using $\varlimsup$ in the definition). The converse is generally false so gain optimality is a weaker criterion than overtaking optimality.

Blackwell (1962), Derman (1962), and subsequent writers have explored the connection between the discounted criterion as $\alpha \uparrow 1$ and the gain criterion. In order to apply Derman (1962) as in Sobel (1970b), suppose X is a finite set and $Y(x)$ is a finite set for each $x \in X$. Then C is a finite set so the set Λ_0 of mappings from X to Y (i.e., x is mapped into $Y(x)$ for each $x \in X$) is finite. Let Λ denote the finite subset of Λ_0 that comprises mappings satisfying (5.12). Thus Λ is finite too.

If $\alpha < 1$, Theorem 5.2(a) asserts the existence of an optimal policy that is a member of Λ. Let $A(x, \alpha)$ denote the dependence on $\alpha < 1$ of such a policy and let $\alpha_1, \alpha_2,...$ satisfy $\alpha_k < 1$ for all k and $\alpha_k \to 1$. Then $A(\cdot, \alpha_k)$ is an element of Λ, a finite set, so the sequence $A(\cdot, \alpha_1)$, $A(\cdot, \alpha_2)$,..., must contain a subsequence all of whose members are the same element of Λ. Let $A(\cdot) \in \Lambda$ denote the policy corresponding to such a subsequence $\alpha_{n(1)}, \alpha_{n(2)},...$, i.e., $A(\cdot) \equiv A(\cdot, \alpha_{n(1)}) \equiv A(\cdot, \alpha_{n(2)}) \equiv \cdots$. Then the argument in Derman (1962) establishes

$$\Gamma(A(\cdot)|x) = \lim_{k \to \infty} (1 - \alpha_{n(k)}) f(x, \alpha_{n(k)}), \qquad x \in X,$$

$$\Gamma(A(\cdot)|x) = \sup_Z \Gamma(Z, x), \qquad x \in X,$$

where $f(x, \alpha)$ makes explicit the dependence of $f(\cdot)$ on the value of $\alpha < 1$. Therefore, $A(\cdot)$ is gain optimal. But $A(\cdot) \in \Lambda$ so $A(\cdot)$ satisfies (7.3).

Let $\mathbf{N}$ denote the set of natural numbers $\{0, 1, 2,...\}$. Then the preceding argument justifies the following claim.

THEOREM 7.1. *If* $X = \{0, 1, 2,..., \sigma\}$ *for some* $\sigma \in \mathbf{N}$ and $Y(x) = \{[0, x] \cap \mathbf{N}\}$, $x \in X$ *then the assumptions of Theorem* (5.2a) *imply existence of a gain-optimal policy that satisfies* (5.12).

It would be interesting to verify Theorem 7.1 without restricting X and $Y(x)$ for each x to be finite sets. If only X is finite, then the problem still seems surprisingly delicate. Suppose in this case there exists

$$A(x) \equiv \lim_{\alpha \uparrow 1} A(x, \alpha), \qquad x \in X. \tag{7.4}$$

Results of Fox (1967) show that $A(\cdot)$ may *not* be gain optimal unless the Markov chain structure induced by $A(\cdot)$ is also the chain structure induced by every one of $A(\cdot, \alpha_k)$, $k = 1, 2,...$, where $\alpha_k \uparrow 1$.

The difficulty of establishing (7.4) is another obstacle to generalizing Theorem 7.1. The difficulty does not seem to stem from finiteness of X. One might conjecture that optimal consumption is a nonincreasing function of α so that $A(x, \cdot)$ is nondecreasing for each $x \in X$. Then the limit in (7.4) would exist because $A(x, \alpha) \in Y(x)$ so $A(x, \alpha) \leqslant x$ from $Y(x) = [0, x]$ and $A(x, \cdot)$

would be a bounded monotone function. To establish monotonicity of $A(x, \cdot)$ one might exploit Theorem 5.2(a) and first establish monotonicity of $A_n(x, \cdot)$ for all n (let the dependence on α of $A_n(x)$ and $f_n(x)$ be explicit). A straightforward inductive proof shows for each n that $f_n(x, \alpha)$ is concave and nondecreasing as a function of α. However, this property does not ensure monotonicity for $A_n(x, \cdot)$. An argument similar to the proof of Theorem 4.2 shows that a sufficient condition would be $J_n^{(3)}(x, y, \alpha)$ nondecreasing in α (for each n, x, and y). In turn, $J_n^{(2)}(x, y, \cdot)$ would be nondecreasing if $f_{n-1}^{(1)}(x, \alpha)$ were a nondecreasing function of α (for each x). This last step has not been accomplished nor is there a counter example.

Bewley (1977) proves, essentially, that $A(x, \cdot)$ is nondecreasing for each $x \in X$ when $X = \mathbf{R}_+$, $Y(x) = [0, x]$, $s(y, d) \equiv y + d$, and $G(x, y) \equiv g(x - y)$ is differentiable and strictly concave. The steps of his proof are similar to those above which we had previously outlined.

ACKNOWLEDGMENTS

We are grateful to Professor Alvin Klevorick for several stimulating discussions and to a referee who noted an error in an earlier version of Theorem 6.1 and suggested expository improvements throughout the paper.

REFERENCES

R. AMIR, Optimum operation of a multi-reservoir water supply system, Ph. D. Thesis, Stanford University (1967).

T. BEWLEY, The permanent income hypothesis: a theoretical formulation, *J. Econ. Theory* **16** (1977), 252–292.

D. BLACKWELL, Discrete dynamic programming, *Ann. Math. Statist.* **36** (1962), 719–726.

E. S. BOYLAN, On the avoidance of extinction in one sector growth models, unpublished manuscript, 1977.

W. A. BROCK AND M. MAJUMDAR, Global asymptotic stability results for multisector models of optimal growth under uncertainty when future utilities are discounted, *J. Econ. Theory* **18** (1978), 225–243.

W. A. BROCK AND L. J. MIRMAN, Optimal economic growth and uncertainty: the discounted case, *J. Econ. Theory* **4** (1972), 479–513.

W. A. BROCK AND L. J. MIRMAN, Optimal economic growth and uncertainty: the no discounting case, *Internat. Econ. Rev.* **14** (1973), 560–572.

E. V. DENARDO AND U. ROTHBLUM, Overtaking optimality for Markov decision chains, *Math. Operations Res.* **4** (1979), 144–152.

C. DERMAN, On sequential decisions and Markov chains, *Management Sci.* **9** (1962), 16–24.

W. FELLER, "An Introduction to Probability Theory and its Applications," Vol. 2, second ed., Wiley New York (1971).

B. FOX, Existence of stationary optimal policies for some Markov renewal programs, *SIAM Rev.* **9** (1967), 573–576.

N. HAKAANSON, Optimal investment and consumption strategies under risk for a class of utility functions, *Econometrica* **38** (1970), 587–607.

N. HAKAANSON AND B. L. MILLER, Compound return mean-variance efficient portfolios never risk ruin, Finance Working Paper No. 8, Inst. of Business and Economic Research, Univ. of Calif., Berkeley (1973).

N. HAKAANSON AND B. L. MILLER, Capital accumulation and production for the firm: optimal dynamic policies, *Management Sci.* **12** (1965), 193–205.

B. JAMISON, Asymptotic behavior of successive iterates of continuous functions under a Markov operator, *J. Math. Anal. Appl.* **9** (1964), 203–214.

S. KARLIN, "Total Positivity," Vol. I, Stanford Univ. Press, Stanford, Calif. (1968).

D. LEVHARI AND T. SRINIVASAN, Optimal savings under uncertainty, *Rev. Econ. Studies* **36** (1969), 153–164.

A. S. MANNE, "Investments for Capacity Expansion-Size, Location and Timephasing," M.I.T. Press, Cambridge, Mass. (1967).

P. MASSÉ, "Les Réserves et la Régulation de l'Avenir dans la Vie Économique," 2 Vols., Hermann, Paris (1964).

B. L. MILLER, Optimal consumption with a stochastic income stream, *Econometrica* **42** (1974), 253–266.

L. J. MIRMAN, Uncertainty and optimal consumption decisions, *Econometrica* **39** (1971), 179–185.

L. J. MIRMAN AND I. ZILCHA, On optimal growth under uncertainty, *J. Econ. Theory* **11** (1975), 329–339.

E. PHELPS, The accumulation of risky capital, *Econometrica* **30** (1962), 729–743.

M. ROSENBLATT, Transition probability operators, *in* "Proceedings 5th Berkeley Symp. Math. Statist. Prob." Vol. II, Part 2 (1967), pp. 473–483.

H. L. ROYDEN, "Real Analysis," Macmillan Co., New York (1963).

J. SCHECHTMAN, An income fluctuation problem, *J. Econ. Theory* **12** (1976), 218–241.

J. SCHECHTMAN AND V. ESCUDERO, Some results on "An Income Fluctuation Problem," *J. Econ. Theory* **16** (1977), 151–166.

M. J. SOBEL, Employment smoothing (capital accumulation) with production for stochastic demand, *Management Sci.* **16** (1970a), 340–349.

M. J. SOBEL, Making short-run changes in production when the employment level is fixed, *Operations Res.* **18** (1970b), 35–51.

M. J. SOBEL, Reservoir management models, *Water Resources Res.* **10** (1975).

D. M. TOPKIS, Minimizing a submodular function on a lattice, *Operations Res.* **26** (1978), 305–321.

W. WHITT, Consumption and investment under uncertainty over an infinite horizon, unpublished manuscript (1975a).

W. WHITT, Consumption and investment under uncertainty: The wealth stochastic process, unpublished manuscript (1975b).

M. E. YAARI, A law of large numbers in the theory of consumer's choice under uncertainty, *J. Econ. Theory* **12** (1976), 202–216.

Printed by the St. Catherine Press Ltd., Tempelhof 37, Bruges, Belgium

From ...C OPTIMIZATION AND MATHEMATICAL ECONOMICS
Edited by Pan-Tai Liu
(Plenum Publishing Corporation, 1980)

5

Stochastic Games, Oligopoly Theory, and Competitive Resource Allocation

MARTIN SHUBIK AND MATTHEW J. SOBEL

1. Introduction

An oligopolistic market is one with only a few firms that supply the commodity being purchased. Oligopoly theory, until recently, evolved without regard to the institutional details encountered in specific markets and without addressing the role played by time. Oligopoly models were treated statically, or at best, conversationally dynamically. However, dynamic oligopoly models have been analyzed with increasing frequency in recent years and some of these analyses are responsive to institutional details.

Here we compare the literature on dynamic models of oligopoly with our intepretation of the objectives of oligopoly theory. We use discrete time-sequential games, sometimes called "stochastic games," as a canonical form in which to discuss the issues. The stochastic game model encompasses many interesting oligopoly models and it seems to offer an appropriate level of generality to address research needs. Incidentally, we do not believe that there is any importance to economic theory associated with the distinction between continuous and discrete time models, i.e., between stochastic games and differential games. In principle, the discus-

MARTIN SHUBIK • Cowles Foundation, Yale University, New Haven, Connecticut.
MATTHEW J. SOBEL • College of Industrial Management, Georgia Institute of Technology, Atlanta, Georgia.

sion could be couched in terms of continuous time models instead of stochastic games.

We believe that the primary objective of oligopoly theory is to provide an understanding of pricing and resource allocation over time in large firms, and the consequent market behavior of such firms. We begin by enumerating some issues involving dynamics that are inherent in this goal. We sketch a general stochastic game model and use the model to address the task of constructing satisfactory dynamic oligopoly models. This effort in part becomes a specification of research needs and opportunities in stochastic game theory and oligopoly theory. At this stage in the development of both subjects, it is useful to identify problems rather than only describe past accomplishments. Section 5 cites specific recent results, and Section 6 presents a blending of problems in optimization and survival.

2. Dynamical Issues in Oligopoly

What are some characteristics of the dynamics of pricing and resource allocation in large firms, and their consequent market behavior? Here, we mention three kinds of problems. First, there are the dynamics of the composition of the set of firms in the market. This "entry and exit" problem in oligopoly is the subject of current research but most analyses have either been static or ignored the multiplicity of "players" in such games. A closely related problem is how to distinguish "competition among the few" from "competition among the many." The modeling issue is how large must a market become in order for gamelike individual behavior to become unimportant. Second, in a given oligopolistic market, why do prices fluctuate as they do? In most markets, the prices fluctuate more slowly than the prices of the factors comprising the inputs in the production process. This phenomenon of "sticky prices" is widely recognized but has hardly been analyzed in a dynamic oligopoly model. Lastly, in some oligopolistic markets, there is one firm that acts as a leader in changing the price level. Why? Why is there price leadership behavior in some markets but not in others? Why might a firm act passively as a follower under some conditions but bolt the pack under other conditions?

Another collection of dynamical issues concerns the role of information in market behavior. How do firms tacitly communicate their objectives, strategies, and threats to one another? How do divisions of a large firm communicate with one another so that their decentralized actions are mutually supportive of the overall goals of the firm. This is the general problem of managerial control. Furthermore, how do accounting conventions affect firm and market behavior? Technically, this question can be posed in terms of alternative aggregations of information. Lastly, what are the effects of imperfections in information, particularly those due to delays in transmission of information? Little progress has been made on a general treatment of this last issue and the prior one has been analyzed in some detail only in static models (Ref. 1).

What are the effects of market size? Most facets of this issue are not particularly dynamical in nature but we should know how to analyze them in dynamic models. As one example, what is the effect on product quality of the number and size of firms that are competing? An issue that is primarily dynamical is the dependence of the number of firms in the market upon the time rate at which information spreads, and vice versa.

Preference structures have been treated somewhat incidentally in oligopoly theory. Important research on intertemporal preference orderings is currently being done (Ref. 2) for models of individual decision-making over time. Comparable investigations of dynamic multiperson decision models have not yet begun. The situation becomes even more complicated if we construct "behavioral models of the firm" (Refs. 3 and 4) that discard the notion of a single monolithic "decision maker" making all the decisions in each firm. The models in "team theory," for example, can be construed as noncooperative games among players having the same preference ordering over outcomes but differing in the information and the actions available to each. We have yet to see an investigation of sequential models of this kind. Lastly, the dynamic oligopoly models analyzed thus far are predicated on a scalar objective such as each firm's discounted operating profit. However, various economists argue that managers in firms behave as if they were maximizing vector objectives. Components other than profit might include rate of growth in sales, number of employees, share of the market, and survival. An important first step has been taken in the analysis of sequential games with vector payoffs (Ref. 5) but this general theory has yet to be applied to a dynamic oligopoly model.

We now turn to some issues of constructing satisfactory dynamic oligopoly models. The canonical form of a general stochastic game will be useful for that purpose. The next section briefly defines a stochastic game and enumerates some notions on the "solution" of such a model.

3. Stochastic Games

Let I be a set of *players*, S a set of *states*, and A_s^i a set of *actions* available to player $i \in I$ when the process is in state $s \in S$. These sets are assumed to be nonempty. The composite action of all the players, when the process is in state s, must be an element of $C_s = \times_{i\in I} A^i_s$. We write $a = (a^i) \in C_s$. An outcome of a stochastic game is a sequence $s_1, a_1, s_2, a_2, \ldots$, where $a_t \in C_{s_t}$ for all t. Let $W = \{(s, a): a \in C_s, s \in S\}$.

The dynamics are determined by the decision rules used by the players to choose their actions and by a collection $\{q(\cdot \mid s, a): (s, a) \in W\}$ of probability measures on B_s, the Borel subsets of S. For any period t and $H \in B_s$, if $s_t = s$ and $a_t = a$ then $q(H \mid s, a)$ is the probability that $s_{t+1} \in H$.

A two-person zero-sum matrix game is a special case of a stochastic game where S is a singleton. It is easy to see that in such a game, in general, one may wish to admit randomized strategies. This complicates the measurability and integrability issues that are already embedded in the one-person stochastic game, namely, the Markov decision process. Our exposition suppresses these issues, which are the subject of some current research on stochastic games. The interested reader should read the fine survey by Parthasarathy and Stern (Ref. 6) and the excellent recent paper by Whitt (Ref. 7).

With the preceding caveat, let π^i denote the set of player i's nonanticipative decision rules (including rules that are history dependent and randomized) for choosing a^i_t, for each t, on the basis of the outcome to date, namely, $s_1, a_1, \ldots, s_{t-1}, a_{t-1}, s_t$. The *stationary* policies are particularly interesting decision rules. Let D^i_s be the set of probability measures on the Borel subsets of A^i_s. An element of D^i_s can be used to choose a randomized action when the game is in state s. Let $\Delta^i = \times_{s\in S} D^i_s$, $\Delta = \times_{i\in I} \Delta^i$, and $\pi = \times_{i\in I} \pi^i$. An element of π is a policy. A policy $\gamma \in \pi$ is stationary if there exists $\delta \in \Delta$ such that $\gamma = (\delta, \delta, \delta, \ldots)$ so $a_t = \delta(s_t)$ for all t. We write $\gamma = \delta^\infty$ in this case. Let Δ^∞ denote the subset of stationary policies in π. Finally, it is convenient to represent any $\gamma \in \pi$ as (γ^i, γ^{-i}), where $\gamma^{-i} \in \prod_{j\neq i} \pi^j$.

There has been some research on the ergodic properties of $\{(s_t, a_t)\}$ induced by stationary policies (cf. Refs. 8 and 9) but most literature concerns real-valued payoff functions. Let $r^i(s, a)$ denote the (expected) immediate reward to player i in any period t if the state s_t is s and the composite action a_t is $a \in C_{s_t}$. Let β_i be player i's single-period discount factor and let

$$\begin{aligned} V^i(\gamma \mid s) &= \sum_{t=1}^{\infty} \beta_i^{t-1} r^i(s_t, a_t) \\ v^i(\gamma \mid s) &= EV^i(\gamma \mid s) \end{aligned} \tag{1}$$

denote the total discounted payoff and its expectation when $\gamma \in \pi$ is the policy and $s_1 = s$ is the initial state. Some literature concerns the average payoff per period rather than the discounted payoff but the latter is more appropriate for oligopoly and other open (or partial equilibrium) economic models.

A policy $\gamma \in \pi$ is said to be an *equilibrium point relative to* $H \subseteq S$ iff

$$v^i(\gamma \mid s) = \sup \{v^i(\rho, \gamma^{-1}) \colon \rho \in \pi^i\}, \qquad s \in H, \quad i \in I. \tag{2}$$

We say simply that γ is an *equilibrium point* if it is an equilibrium point relative to S. An equilibrium point relative to H is noncollusively optimal for every initial state in H and every player. Shapley (Ref. 10), in a magnificent early paper, established existence of an equilibrium point amongst stationary policies in a two-player model with $\bigcup_{s \in S} A_s^i$ being a finite set for each player and $r^1(\cdot,\cdot) + r^2(\cdot,\cdot) \equiv 0$. More general existence results concern nonzero sum games played by more than two players (Refs. 6, 7, 11, 12). Also, Henig (Ref. 5) has recently established the existence of an equilibrium point for games where $r^i(\cdot,\cdot)$ is vector-valued.

4. Modeling Issues

We have observed that the extent theory fails to explain why price leadership occurs and which firms are likely to be followers while others are leaders. More generally, there is no satisfactory sequential analogue to "cooperative" theory for static games. The primitive element in most of the cooperative static theory is the coalition, namely, a subset of the players who join together for mutual benefit. However, a satisfactory dynamic theory must admit changes in coalition composition as time passes, but the present theory does not include this feature.

Another modeling issue stems from the embarrassment of riches provided by the size of the set of equilibrium points. It is known that the size of this set increases as information conditions in a game proliferate (cf. Ref. 13). Therefore, oligopoly models that strive to include the design of information systems and managerial control may induce distressingly many equilibrium points. The problem is to decide which one, more generally which subset, is the appropriate object for analysis. We believe that the solution to this problem should vary depending upon the context which motivates the model. In other words, behavioral considerations and institutional details should direct our definition of "the appropriate object for analysis."

It has already been mentioned that no satisfactory canonical model exists to analyze the effects of imperfections in information due to delays in transmission. Indeed, the intricacy of the analysis in a relatively simple case analyzed by Scarf and Shapley (Ref. 14) is alarming. We doubt that a Bayesian approach is appropriate here although one of us has explored this elsewhere (Ref. 8). The extant theory of stochastic games would oblige us to assume that each firm knows the prior distribution held by every other firm.

Careful modeling of many industries leads to the explicit inclusion of bankruptcy conditions in a model. Such conditions exemplify "exit fees" in the class of Markov decision processes called *stopping problems*. There are several interesting qualitative results concerning the structure of optimal policies in stopping problems. As yet, there is no comparable theory for "stopped sequential games." The development of such a theory might have a payoff for oligopoly theory by offering a deeper understanding of the effects of alternative bankruptcy laws and the dynamics associated with the entry and exit of firms from an industry. One of us has suggested a class of "games of economic survival" to pick up the ruin posibilities (cf. Section 6; Refs 15 and 16).

This list of modeling issues is necessarily brief and we have not discussed some pertinent material. Aumann (Ref. 17) has developed results for "supergames." A supergame is a sequence of static games in which the nature of the static games is not contingent on players' past actions. The case of the same static game at each point (in the sequence) has been investigated more than any other. This case is a stochastic game with $|S| = 1$. Friedman (Ref. 18) analyzes this case in oligopoly models. He focuses on "reaction function" strategies; each player's present decision is contingent on the opponent's preceding decision. Such decision rules induce a stochastic game in which S is the set of player's possible single-game decisions. Recently, Rosenthal (Ref. 19) has investigated sequences of games with varying opponents. His point of view may be useful for construction of dynamic models of entry and exit in oligopoly. Shefrin (Ref. 20) has interesting results for dynamic market games with incomplete information.

5. Specific Results

Stochastic game models of oligopoly, even with the limitations enumerated above, are forbiddingly complex. Nevertheless, some progress has been made either by reducing the potential complexity or by building a model for a particular kind of industry and then posing correspondingly special questions. First, we discuss the reduction of complexity.

Stochastic game models are difficult to analyze because the number of players is greater than one, so the players interact with one another, and because the game process extends over time and each player indulges in a variety of intertemporal tradeoffs. Several writers (see Ref. 21 and references therein) have suppressed the complexity owing to the interaction of firms by analyzing models of leader–follower behavior where the identities of the leader and followers are known at the outset. The problem is then the selection of an optimal dynamic policy by the leader and this latter problem is a (one-person) Markov decision process which is much less complex than a stochastic game.

Another suppression of complexity has been obtained by preserving a multiplicity of players (firms) but reducing the original dynamic game to a static game. Specifically, an equilibrium point of a stochastic game is said to be *myopic* if it consists of the *ad infinitum* repetition of an equilibrium point of a static game. The principal sufficient conditions (Ref. 22), satisfied by various dynamic oligopoly models, are:

> for each $i \in I$ and $(s, a) \in W$, $r^i(s, a)$ depends additively on the state s and action a, i.e., there are functions K^i and L^i such that
>
> $$r^i(s, a) = K^i(a) + L^i(s); \tag{3a}$$
>
> transition probabilities depend on the actions taken but not on the state from which transition occurs, i.e.,
>
> $$q(H \mid s, a) = p(H \mid a) \qquad \text{for all} \quad H \in B_S \quad \text{and} \quad (s, a) \in W; \tag{3b}$$
>
> suppose the static game Γ, defined below, has an equilibrium point a_* in pure strategies and let $S^* = \{s : s \in S, a^* \in C_s\}$. Then $p(H \mid a^*) = 1$ for all $H \in B_s$ having $S^* \subseteq H$, i.e., if $s_1 \in S^*$ then a^* is repeatable *ad infinitum* (with probability one). (3c)

Let $\xi(a)$ be a random variable with the measure $p(\cdot \mid a)$ in (3b). Then I is the set of players in the static Nash game Γ, player i's payoff function is $K^i(a) + \beta_i E[L^i(\xi(a))]$, $a \in \bigcup_{i \in I} A^i$, and player i has available the set of moves $A^i = \bigcup_{s \in S} A^i_s$. If a_* is randomized then there is an assumption comparable to (3c). It follows from (3) that $a_t = a_*$ for all t is an equilibrium point [in the stochastic game sense of (2)] relative to S^*.

Numerous Markov decision processes in the literature satisfy (3) but the myopia of their optimal policies was either overlooked or deduced by special and sometimes intricate arguments. Also, various oligopoly models satisfy (3). Kirman and Sobel (Ref. 23) assume that firms make production and pricing decisions each period and that they have linear production costs and arbitrary single-period inventory-related costs. The oligopoly model in Sobel (Ref. 24) focuses on advertising decisions. It is assumed there that each firm's demand each period is a random variable whose distribution depends on all firms' "goodwill." The goodwill is an exponentially weighted moving average of past amounts spent on advertising. Myopia has been applied to other oligopoly models where the competition involves expenditures on research and development, expansion of capacity, and the harvesting of interacting fish species in a coastal fishing industry (Ref. 25).

6. Games of Economic or Social Survival

The goals of profit maximization or cost minimization are present in many economic models. Games of survival, as characterized by Shubik (Ref. 15), stress the binary outcomes of survival or nonsurvival. Yet in many social, ecological, and economic processes the goals include both survival and optimization of the quality of life for the survivor.

Below we describe a general class of games of social or economic survival. With notational changes, they can be recast as stochastic games.

An n-person game of economic survival is described as follows by

$$\alpha^i(t);\ W_1^i;\ B_1^i;\ S_1^i;\ L_1^i;\ \Psi_1^i;\quad \text{and}\quad V^i,$$

where

$\alpha^i(t)$, $i = 1, \ldots, n$, are the single-period *payoffs* faced by the players at time t. The payoffs depend on actions described below and will in general depend upon time.

W_t^i, $i = 1, \ldots, n$ and $t = 1, 2, \ldots$, are the *wealths* of the players at the start of time t. Initial assets W_1^1 and W_1^2 are given as parameters.

B^i are the *ruin conditions*, bankruptcy levels or "absorbing barriers," for the player; i.e., if the assets (or strength or viability) of a player i drop to below B^i, that individual is out of the game.

S^i are the *survival values*; i.e., if individual i is the sole survivor in the game, S^i is the present value of the remaining one-person game.

L^i are the *liquidation values*. If an individual i is ruined, he may still have residual assets at the point of ruin. The value of these assets is given by L^i.

Ψ^i are *discount factors*. We assume that each individual has a discount factor on future consumption. The Ψ^i could be dependent upon the age of the individual, thus reflecting life-cycle considerations.

T is the *time* at which bankruptcy (if any) occurs:

$$T = \inf\{t: (W_t^1, W_t^2) \not> (B^1, B^2)\}.$$

V^i is the *payoff function* to player i. It cannot be fully specified until the strategies and the relationship among income, consumption, and survival are specified.

Two versions are given, the game where pure survival is the goal and the game where the maximization of expected discounted consumption [or utility $\phi^i(\cdot)$ of consumption] is optimized.

At the start of any time t an individual i has W_t^i. A Markov strategy by an individual i is a plan for the selection of an investment amount x_t^i and a consumption amount b_t^i dependent upon $W_t = (W_t^1, \ldots, W_t^n)$.

$$W_{t+1}^i = \xi_t[W_t^i + \alpha^i(x_t^1, \ldots, x_t^n) - b_t^i - x_t^i],$$

where $0 < x_t^i + b_t^i < W_t^i$ and $\xi_1, \xi_2, \ldots$ are independent and identically distributed random variables.

For the game of *pure survival*,

$$V^i = \begin{cases} 1 & \text{if } W_t^i > B^i \quad \text{for} \quad t = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

For the game of *pure consumption optimization*,

$$V^i = \begin{cases} \sum_{t=1}^{\infty} (\Psi^i)^{t-1}\phi^i(b_t^i), & W_t^i > B^i, t = 1, \ldots, \infty, \\ & \text{and game continues indefinitely} \\ \sum_{t=0}^{T} (\Psi^i)^{t-1}\phi^i(b_t^i) + (\Psi^i)^T S^i, & W_t^i > B^i, t = 1, \ldots, T, \\ & \text{and game continues until } T \\ \sum_{t=1}^{T} (\Psi^i)^{t-1}\phi^i(b_t^i) + (\Psi^i)^T L^i, & W_t^i > B^i, t = 1, \ldots, T-1, \\ & W_T^i \le B^k, \text{ and game continues} \\ & \text{until } T-1. \end{cases}$$

For either criterion, let J indicate which player (if either) is ruined:

$$J = \begin{cases} 0 & \text{if} \quad T = \infty, \\ j & \text{if} \quad W_T^j \le B^j. \end{cases}$$

Then, for either criterion, consider the vector (V^1, V^2, T, J). The adoption of a policy by the players induces a (joint) probability distribution of this vector and we may compare the distributions induced by alternative policies. An obvious comparison is according to $v^i = E(V^i)$. For example, in the pure survival criterion, $v^i = P\{V^i = 1\} = P\{J \neq i\}$. Although we shall not pursue the matter here, the two criteria can be treated in a unified manner by first defining an appropriate stochastic game.

6.1. A "Guns or Butter" Example

A two-player example illustrates the tradeoff between consumption and survival. Let b_t and x_t denote player 1's consumption and investment in period t and let d_t and y_t denote the same amounts for player 2.

Let

$$\alpha^1(t) = \begin{cases} A\dfrac{x_t}{x_t + y_t} & \text{if} \quad x_t + y_t > 0, \\ 0 & \text{if} \quad x_t = y_t = 0, \end{cases}$$

$$\alpha^2(t) = \begin{cases} A\dfrac{y_t}{x_t + y_t} & \text{if} \quad x_t + y_t > 0, \\ 0 & \text{if} \quad x_t = y_t = 0. \end{cases}$$

Let $W_1^1 = M$; $W_1^2 = m$; $B^1 = B^2 = L^1 = L^2 = 0$, $S^1 = S^2 = A/(1-\Psi)$; $\Psi^i(c) = c$; $\Psi^1 = \Psi^2 = \Psi$, where $0 < \Psi < 1$. Say $b_t^1 = b_t$ and $b_t^2 = d_t$.

If the goals are pure survival then $x_t = y_t = b_t = d_t = 0$ gives $M_{t+1} = \xi_t M_t > 0$, $m_{t+1} = \xi_t m_t > 0$, where $\xi_t > 0$ is a random variable.

If the goals are maximization of expected consumption then if a solution with joint survival is feasible, player 1 wishes to

$$\max_{b_t, x_t} E \sum_{t=1}^{\infty} (\Psi^i)^{t-1} b_t, \qquad b_t \geq 0, \qquad x_t \geq 0$$

subject to $0 \leq b_t$, $0 \leq x_t$,

$$b_t + x_t \leq M_t$$

and

$$M_{t+1} = \xi_t \left[M_t - b_t + A\left(\frac{x_t}{x_t + y_t}\right) - x_t \right]$$

and player 2 wishes to

$$\max_{d_t, y_t} E \sum_{t=1}^{\infty} (\Psi^i)^{t-1} d_t$$

subject to $0 \leq d_t$, $0 \leq y_t$,

$$d_t + y_t < m_t$$

and

$$m_{t+1} = \xi_t \left[m_t - d_t + A\left(\frac{y_t}{x_t + y_t}\right) - y_t \right].$$

Let $\mu = E[\xi_t]$. If $\Psi\mu \leq 1$ and $P\{\xi \geq \Psi\mu/2\} = 1$ (for which $P\{\xi \geq \frac{1}{2}\} = 1$ is sufficient) then we can show that there is a myopic equilibrium point with respect to $\{(M_1, m_1): (M_1, m_1) \geq \Psi\mu A/4(1, 1)\}$ given by

$$\begin{aligned} x_t &= y_t = \mu A\Psi/4, && \text{for } t = 1, 2, \ldots, \\ b_t &= d_t = A(2\xi_{t-1} - \Psi\mu)/4, && \text{for } t = 2, 3, \ldots, \\ b_1 &= M_1 - \mu A\Psi/4, && d_1 = m_1 - \mu a\Psi/4 \end{aligned}$$

and expected payoffs are

$$\begin{aligned} v^1 &= M_1 - A\mu\Psi/4 + \sum_{t=2}^{\infty} \Psi^{t-1}(2\xi_{t-1} - \Psi\mu)A/4 \\ &= M_1 + A\mu\Psi/[4(1 - \Psi)], \\ v^2 &= m_1 + A\mu\Psi/[4(1 - \Psi)]. \end{aligned}$$

If $P\{\xi_1 < \Psi\mu/2\} > 0$, or $(M_1, m_1) \ngeq \Psi\mu A/4(1, 1)$, or $\Psi\mu > 1$ then the analysis and solution become complicated.

The "net earnings" or gains from competition are given by the terms

$$A\frac{x_t}{x_t + y_t} - x_t.$$

These portray the resource struggle or the "battle conditions." The payoffs involve only the b_t, i.e., the resources drawn off for consumption.

ACKNOWLEDGMENT

The preparation of this work was partially supported by NSF grants ENG76-46156 and SOC78-T5730. This work relates to Department of the Navy Contract N00014-77-C-0518 issued by the Office of Naval Research under Contract Authority NR-047-006. However, the content does not necessarily reflect the position or the policy of the Department of the Navy or the Government, and no official endorsement should be inferred.

References

1. MARSCHAK, J. and RADNER, R., *Economic Theory of Teams*, Yale University Press, New Haven, 1972.
2. KREPS, D. M. and PORTEUS, E. L., Temporal resolution of uncertainty and dynamic choice theory, *Econometrica*, Vol. 46, pp. 185–200, 1978.
3. CYERT, R. M. and MARCH, J. G., *A Behavioral Theory of the Firm*, Prentice Hall, Englewood Cliffs, N.J., 1963.
4. WILLIAMSON, O. E., *Markets and Hierarchies*, Free Press, New York, 1975.
5. Henig, M., *Multicriteria Dynamic Programming*, Yale University, Ph.D. Thesis, 1978.
6. PARTHASARATHY, T. and STERN, M., On Markov games, *Differential Games and Control Theory*, Vol. 2, Edited by E. Roxin, P. T. Liu, and R. Sternberg, Marcel Dekker, New York, pp. 1–46, 1977.
7. WHITT, W. Representation and approximation of noncooperative stochastic games," *J. Siam*, to appear.
8. SANGHVI, A. P. and SOBEL, M. J., Bayesian games as stochastic processes, *International Journal of Game Theory*, Vol. 5, pp. 1–22, 1976.
9. SANGHVI, A. P., Sequential games as stochastic processes, *Stochastic Processes and Their Applications*, Vol. 6, pp. 323–336, 1978.
10. SHAPLEY, L. S., Stochastic games, *Proceedings of the National Academy of Sciences, USA*, Vol. 39, pp. 1095–1100, 1953.
11. ROGERS, P. D., *Nonzero-Sum Stochastic Games*, Operations Research Center, University of California, Berkeley, Report ORC-69-8, 1969.
12. SOBEL, M. J., Noncooperative stochastic games, *Annals of Mathematical Statistics*, Vol. 42, pp. 1930–1935, 1971.
13. DUBEY, P. and SHUBIK, M., Information conditions, communications, and general equilibrium, *Mathematics of Operations Research*, 1979, in press.
14. SCARF, H. E. and SHAPLEY, L. S., Games with partial information, *Contributions to the Theory of Games*, Volume III, Edited by M. Dresher, A. W. Tucker, and P. Wolfe, Princeton University Press, Princeton, N.J., pp. 213–230, 1957.
15. SHUBIK, M., *Strategy and Market Structure*, Wiley, New York, 1958.

16. SHUBIK, M. and THOMPSON, G. L., Games of economic survival, *Naval Research Logistics Quarterly*, Vol. 6, pp. 111–123, 1959.
17. AUMANN, R. J., Acceptable points in general cooperative n-person games, *Contributions to the Theory of Games. IV*, Edited by R. D. Luce and A. W. Tucker, Annals of Mathematics Study 40, Princeton University Press, Princeton, New Jersey, pp. 287–324, 1959.
18. FRIEDMAN, J. W., *Oligopoly and the Theory of Games*, North-Holland, Amsterdam, 1977.
19. ROSENTHAL, R. W., Sequences of games with varying opponents, *Econometrica*, 1979, in press.
20. SHEFRIN, H. M., Bayesian convergence in market games with incomplete information, *Proceedings of the Third Kingston Conference on Differential Games and Control Theory, 1978*, Marcel Dekker, to appear.
21 LIPPMAN, S. A., Optimal pricing to retard entry, *Review of Economic Studies*, Vol. 44, p. 32, 1977.
22. SOBEL, M. J., Myopic equilibrium points in stochastic games, unpublished manuscript, 1978.
23. KIRMAN, A. P. and SOBEL, M. J., Dynamic oligopoly with inventories, *Econometrica*, Vol. 42, 279–287, 1974.
24. SOBEL, M. J., "Sequential Marketing Games," unpublished manuscript, 1977.
25. SOBEL, M. J., "A Stochastic Game Model of a Regional Fishery," unpublished manuscript, 1979.

Part II

Dynamical Theory of the Economics of Extractive Natural Resources

Naturally growing extractive resources are renewable or nonrenewable, depending on whether they are regenerated at significant rates over time. Examples of renewable resources are fisheries, forests, and ground water; of nonrenewable resources, hard minerals and the hydrocarbon fuels.

While Hotelling applied the calculus of variation to the nonrenewable resource problem as early as 1931, systematic though simplistic treatment of the inherently dynamic model problems did not begin until the late 1950s. Recent years have witnessed a flourishing body of economic literature applying modern control theory to extractive resources.

The recent energy crisis and the extension of jurisdiction over marine fish stocks have given rise to numerous concerns in the public sector. Principal among the concerns are the resource scarcity and exploration, production efficiency and distribution under alternative market structures and institutional arrangements, and optimal policies to achieve various social goals. The five papers in this part treat various aspects of these important issues.

The essence of the extractive resource problem is a differential equation of the form

$$\dot{x} = F(x) - h(t), \tag{1}$$

where $x(t)$ is the size of resource stock at time t, $F(\cdot)$ is the rate of natural growth, and $h(t)$ is the rate of extraction. For nonrenewable resources, the growth rate is negligible and $F(\cdot) = 0$. The typical problem is to optimize some functional subject to (1) and examine the intertemporal path of $h(t)$ and possibly other control variables.

As Clark has pointed out, harvesting of renewable resources is a problem of managing a flow of goods and can be studied on a capital-theoretic framework. In this respect, the problem is analogous to that of economic growth. Stability of a steady-state solution is often one of the most crucial points in the analysis of such a problem, because it pertains to the question of whether a species could be sustained at a certain level under some harvesting policy. Clark's book on optimal management of renewable resources contains a comprehensive treatment of many aspects of the problem.

In this part, Clark extends his results on optimal harvesting policies for sole ownership of a fishery resource property to the case of multi-ownership of a common property. The problem is formulated as an N-person nonzero-sum differential game. Under the assumption that competition exists among all agents, a Nash equilibrium solution is obtained. With such a solution, the more "efficient" agent may be able to eliminate his competitors if the level of stock is reduced to an appropriate level. He may even be able to operate at the optimum level of sole ownership if the effective advantage is large. A more general model of restricting each agent's input level is also discussed. With relatively simple models and mathematical analysis, this paper presents and discusses many economic issues in the exploitation of common property fishery resources by using the game-theoretic approach.

For an exhaustible resource, the differential equation describing the process of exploitation is simpler, since $F(\cdot) = 0$. But despite the simplicity of the equation, the problem is often complicated by many factors. For one thing, the amount of reserve is usually not exactly known and remains to be determined as exploitation goes on. This requires stochastic and adaptive control techniques. When stochastic models are used, the firm's policies are very much dependent on its attitude toward risk (risk taking or risk averse) and the degree of uncertainties as measured by some statistical parameters. The mathematical complexities increase as more economic aspects are taken into consideration. Basically, this is a bounded state control problem in which the steady-state solution no longer has the significance of a dynamic equilibrium; rather, it yields the final levels of accumulated extractions and determines whether the resource will be depleted.

In Liu's paper, the problem of exploitation of exhaustible resources is tackled from a mathematical point of view. Some techniques for obtaining the optimum extraction rate for a firm over an infinite horizon are discussed. A general profit function and cost function associated with extraction are assumed. Uncertainty in the amount of reserve is also considered. He shows that at a fixed amount of accumulated extraction, the optimum extraction rate is always lower with uncertainty than without uncertainty. A significant feature of the analysis is the use of the transversality condition at infinity, from which the final level of accumulated extraction can be determined. Moreover, for n-firm exploitation, the transversality condition leads to an equilibrium condition among n firms under Pareto optimality. Different levels of accumulated extraction for different firms can also be obtained.

Arrow and Chang employ a model that includes both exploration and consumption of uncertain natural resources. The resource is assumed to be Poisson distributed throughout a relevant area. The amount of (unknown) reserve and the unexplored area at any time are the state variables, while the rate of consumption and the exploratory effort are the control variables. The economy seeks to maximize the integral of a utility function minus the cost of exploration. This is an optimal control problem with a jump process. Using the method of dynamic programming, they derive the equivalent of Bellman's equation. From that equation, the optimal policy and the optimum return function are characterized. The optimum exploratory effort almost always alternates between zero and infinity, and the impact of such an alternation is examined.

Lewis and Schmalensee investigate the implication of two intermediate market structures for supplier behavior. The first is a market dominated by a cartel that maximizes its profit subject to the price-taking behavior of a competitive fringe of many suppliers. An interesting result of their analysis is that when allowed to misrepresent the true extent of its reserve holdings, the cartel's optimal policy is not to lie but to tell the truth.

Lewis and Schmalensee also present a model of a Cournot–Nash oligopoly in which each firm is large enough to have some control over price. Existence and uniqueness of an equilibrium are established, and some comparative dynamic results are obtained under conditions of perfect information and homogeneous reserve holdings and extraction costs.

It has been long recognized in resource economics that risk taking plays an important role in a competitive market and affects resource allocation. Sutinen examines the implications of royalties for production over time. He notes that producers of petroleum and other exhaustible resources typically do not own the resource but rather purchase a lease giving them the right to extract the resource. A common feature of such leases is payment to the owner a royalty plus a fixed amount of payment.

Sutinen constructs a competitive model of the market for extractive rights, with resource owners being the suppliers and the producing firm the demander of the rights. Demand by consumers of the final product is assumed stochastic in all periods. Under these conditions, two sets of leasing arrangements are analyzed: a fixed rental and a class of risk-sharing arrangements involving royalty payments. The resulting sequence of production is characterized for each set of leasing arrangements and compared with the optimal production sequence. When both owners and firms are risk averse, Sutinen finds that royalties yield a production sequence superior to that of a fixed rental lease.

Reprinted from JOURNAL OF ECONOMIC THEORY
All Rights Reserved by Academic Press, New York and London
Vol. 22, No. 1, February 1980
Printed in Belgium

Incentives, Iterative Communication, and Organizational Control

SUSAN I. COHEN*

College of Industrial Management, Georgia Institute of Technology, Atlanta, Georgia 30332

Received March 22, 1978; revised April 10, 1979

I. INTRODUCTION

This paper defines a class of incentive mechanisms for decentralized organizations when communication between the members of the organization must, a priori, be iterative in nature. Payoff functions for the members of the organization are defined and these functions along with the set of possible member strategies are viewed as defining an N-person game. We define a solution concept for this game that has the following properties: the members of the organization (nontrivially) maximize their payoffs at a "solution" *and* the overall organizational goals are achieved by any "solution."

Acting in accordance with the behavioral rules of the organization, i.e., "telling the truth," is shown to be among the "solutions" to the game. Thus, the members of the organization are shown to have an incentive to follow these rules, since doing so is an individually optimal strategy that is, in some sense, easy to calculate.

Arrow [1] has defined the problem of organizational control as consisting of two parts: (1) the choice of operating (behavioral) rules, i.e., communication and decision making rules and (2) the choice of enforcement rules, i.e., rules that induce the members of the organization to follow the operating rules.[1] There is a vast literature in economics concerned with the definition and/or analysis of operating rules for decentralized organizations, theoretical planning procedures being the most obvious [2, 14, 18].[2] These various

* This paper contains results from my Ph.D. thesis submitted to Northwestern University. I want to thank my advisor Theodore Groves for his encouragement and help. This paper has benefited greatly from comments by Matthew Sobel, David Nachman, and an anonymous referee. All errors, of course, remain my own.

[1] Obviously, the enforcement rules defined for a particular organization must be tied to individual members' preferences, either directly or indirectly.

[2] The mathematical programming literature contains many different algorithmic procedures whose economic interpretations are as decentralized planning procedures: e.g., the Dantzig–Wolfe decomposition algorithm [4]. In addition many economic planning procedures are based on mathematical programming algorithms: for example, Malinvaud's

0022-0531/80/010037–19$02.00/0

literatures are concerned with normative analyses: Do procedures exist that have certain desirable properties.·[3] The procedures defined in the literature have several important characteristics in common: they inevitably involve iterative communication and they generally involve decentralized decision making, due to the size of the problems considered. None of these procedures deal with the provision of incentives (the definition of enforcement rules) to ensure that the operating rules are followed; they thus ignore the second part of the organizational design question.

An important question is thus raised: Are there enforcement rules that make compliance with the behavioral rules in the best interests of the members of the organization.· If the answer to this question is no, then it is not necessarily true that these procedures will lead to a solution of the organization's overall problem.

Until recently, very little work has been done on the specification of the enforcement rules, referred to in other contexts as the incentive problem. The existing work has examined the properties of classes of incentive mechanisms [8, 13], investigated incentives in a team [9], and solved the "free rider" problem in a general equilibrium model with both public and private goods [12]. In all of this work, communication is noniterative in nature.[4]

The model and analyses developed in this paper are in the same spirit as the model and analyses of Groves [10, 11] and Groves and Loeb [13]. These latter models will be referred to as "Groves schemes." Implementation of a Groves scheme requires each member of the organization to have sufficient computational, storage (memory), and communcation capabilities both to determine the graph of the function it must transmit to the central decision maker and to make this transmission (without error) as a single message. The informational capabilities of most economic decision making units are severely limited in reality. The purpose of this paper is the exploration of the economic implications of these limitations on the production and communication of information.

As a first step in this exploration, this paper examines the incentive properties of a particular (rather general) scheme in which communication is iterative. This scheme implicitly reflects some limitations on computational capacity, memory, and communication capabilities of the members of the organization. It is only necessary that the members be able to calculate the

"Procedure Implying Mathematical Programming at the Center" [18] is an inner-linearization algorithm for a concave programming problem. For a discussion of some of the economic procedures, see [14]; of the mathematical programming algorithms, see [6].

[3] "Nice" properties may be of the following type: efficient use of information, monotonicity, feasibility, etc. (see [14]).

[4] Groves and Loeb [13] use an iterative procedure to reconstruct functions (at least locally) by a Taylor series expansion. Dreze and de la Vallee Poussin [5] have dealt with incentives in a tâtonnement-type procedure.

value of the function of interest at a particular point in its domain, and communicate an associated message. The iterative nature of the communication process implicitly captures the fact that the members of the organization can do no more than that without incurring real economic or time costs that may change the nature of the function itself. This would change the nature of the organization's problem, since information production and communication efforts would have to be considered explicitly.

The results obtained in this paper can be applied to establish the incentive properties of some decentralized planning procedures. Malinvaud's [18] "Procedure Implying Mathematical Programming at the Center" is a case in point. The firms in the economy compute and transmit, to a central authority, well-chosen points in their production possibilities sets in response to well-chosen price vectors sent by the central authority. It is easily shown that this procedure is a special case of the procedure developed in this paper (see Cohen [3]). While it is straightforwqrd to contract Malinvaud's procedure into a Groves scheme, this action would require the firms to have a communicable description of their entire production possibilities sets. In practice, however, knowledge of production possibilities is very often limited to the ability to determine outputs or an input–output vector for a given vector of prices. Hence, the use of an iterative scheme becomes feasible as well as prudent. Over an extended period of time, assuming fixed technology, it is possible that some firms may have more or less complete knowledge of some portions of their production possibilities sets. However, for a Groves scheme to be implemented, every firm must have complete, communicable knowledge of all of its production possibilities set.

That the consequences of these informational limitations are not trivial is demonstrated by a comparison of the properties of a Groves scheme and the iterative scheme developed here. The existence of dominant strategy equilibria can be shown for a Groves scheme. In addition, "telling the truth" is one such equilibrium. Even though the iterative scheme developed in this paper is under certain circumstances informationally equivalent to a Groves scheme, one must settle for weaker results.

In Section II, a model of a decentralized organization is defined. Strictly for ease of exposition, the analysis in this paper is couched in terms of a multidivisional firm that seeks to maximize overall firm profit.

Given the iterative operating rules, the organizational design problem is defined in Section III as one of finding enforcement rules that lead to compliance with the operating rules. We then define (in Section IV) the problem for any member of the organization as the choice of messages to be sent to the central decision maker (the Center) at each iteration of the communication process, i.e., the choice of a strategy. The operating and enforcement rules and the set of possible member strategies are viewed as defining an N-person (non-zero sum) game in normal form.

It is assumed that this game is played noncooperatively. There are two possible ways of viewing the game. It could be viewed as a game in which payoff can be received only when the communication process has been run to its end and terminal decisions have been taken and/or implemented. This formulation does not allow us to deal with truncated communication and suboptimal decision making. Alternatively, one could view the game as one in which decisions can be taken and payoff recieved at any iteration.

As a first attempt at an analysis of incentives when communication is iterative, we view the game as being of the first type. A natural solution concept for such an N-person noncooperative game is that of the Nash equilibrium. From a normative, or design, point of view, just knowing that a Nash equilibrium exists is not enough; in addition, one would like to know whether such an equilibrium will be attained. Thus, we might ask for the existence of a dominant strategy equilibrium that is, in some sense, easy to calculate. As was noted above, the iterative nature of communication precludes demonstrating the existence of such equilibria for the present game. In Section V, a "solution" to the game is defined as a Nash equilibrium that weakly dominates in the class of Nash equilibria; this definition avoids some of the possible ambiguities of multiple Nash points.

In Section VI, a general class of operating rules is defined, and we demonstrate the existence of enforcement rules with certain desirable properties. It is shown that any "solution" for the game, defined by these operating and enforcement rules, leads to a decision solving the organization's overall problem. In addition, a dominant strategy equilibrium is shown to exist for a restricted set of player strategies. Moreover, the equivalence class of "truthful" strategies is shown to be a subset of the class of restricted dominant strategy equilibria, as well as a subset of the set of "solutions."

Finally, in Section VII, we discuss the applicability of our operating rules to existing decentralized planning procedures and mathematical programming algorithms.

II. A Model of a Multidivisional Firm with N-Divisions

Consider a multidivisional firm with N divisions and a corporate center (the Center). Each division has a profit contribution function $\Pi_i(\cdot)$, where $\Pi_i(x)$ is the profit contribution by division i given the decisions x.[5] The Center's direct contribution to profit is denoted by the function $\Pi_0(\cdot)$.[6]

[5] It is not specified, as yet, who makes the decisions.

[6] It has been pointed out that inclusion of the Center's contribution function does not preclude the application of a Groves (noniterative) mechanism, and that is indeed the case. The purpose of the present work is to investigate the incentive properties of mechanisms in which communication must be iterative. The $\Pi_0(\cdot)$ function is included because, especially in the context of the multidivisional firm, one can envision situations in which the Center contributes to overall profit directly.

The L-dimensional vector x may represent an input vector of resources or x may be the level of a "public good" used by all the divisions and the Center. The total profit of the firm is the sum of each division's, and the Center's, profit contribution; i.e., total profit $= \sum_{i=1}^{N} \Pi_i(\cdot) + \Pi_0(\cdot)$. The Center's problem is to choose the vector x, subject to constraints, to maximize the firm's profit. Formally, the Center's problem is assumed to be to solve

$$\text{(A)} \qquad \operatorname*{Max}_{x} \sum_{i=1}^{N} \Pi_i(x) + \Pi_0(x) \qquad \text{s.t.} \quad x \in X$$

where $\Pi_i : \mathbb{R}^L \to \mathbb{R}$ is concave for all $i = 0, 1,..., N$ and X is a convex, compact set. More realistic models of decentralized organizations would permit nonconvexities in X and departures from concavity in the $\{\Pi_i(\cdot)\}$ but (A) is a reasonable starting point to develop a theory of control with iterative communication.

Suppose that each division of the firm knows[7] its own profit contribution function $\Pi_i(\cdot)$, but not the functions of the other divisions. Also, suppose that these functions are not known by the Center either; however, it does know that they are concave. Only the Center knows that function $\Pi_0(\cdot)$ and the set X.

Since the Center does not know the divisions' $\Pi_i(\cdot)$ functions, it must seek some information from the divisions about them. Once the Center has this information, it must make a decision, i.e., choose a "good" x vector.

III. The Control Mechanism

A *control mechanism* is defined by a complete specification of both operating and enforcement rules. These rules are known explicitly by the Center. For the iterative procedures considered here, the operating rules consist of a message set for the divisions, a message rule for the Center for each possible iteration, and a decision rule for the Center.[8] Thus, an *iterative control mechanism* for problem (A) is defined by a quadruple: The first component is the message set for the divisions; that is, a set from which a division at each iteration may select a message to be sent to the Center. The second component is a sequence of rules for messages sent by the Center, specifying at every iteration the message to be sent to the divisions given the messages received from the divisions at all previous iterations. The third component is the Center's decision rule. Given the messages

[7] Where "know" means that for each x, division i can calculate $\Pi_i(x)$.

[8] For a discussion of control mechanisms for noniterative procedures, see Groves [10]

received from the divisions at all iterations, this rule specifies the Center's decision. The final component of the quadruple is an N-tuple of evaluation measures, one for each division. Each division's evaluation, as specified by these measures, will depend on its realized contribution Π_i, which we assume can be observed ex post by the Center, and on the messages of all the divisions. This last component of the quadruple implicitly specifies the enforcement rules for the control mechanism; i.e., maximization of a division's evaluation measure enforces the use of the prescribed operating rules.

Formally, an *iterative control mechanism* C is defined by

$$C \equiv (\mathscr{M}, \{y^t(\cdot)\}_{t=1}^{\infty}, x(\cdot), \langle E_i(\cdot)\rangle_{i=1}^{N}), \tag{1}$$

where

(a) $\mathscr{M} \subset$ {finite-dimensional Euclidean space} is the *message set of the divisions.* $m_{it} \in \mathscr{M}$ is the message of division i at iteration t.

$$m_t \equiv (m_{1t}, ..., m_{Nt}),$$
$$m^t \equiv (m_1, ..., m_t),$$
$$m_i^t \equiv m_{i1}, ..., m_{it}).$$

(b) y^t: $\mathscr{M}^{N(t-1)} \to Y$ is the *message rule of the Center* for $t \geqslant 2$. y^1 is a given, fixed element of Y.

(c) x: $\mathscr{M}^{\infty} \to X$ is the *Center's decision rule.*[9]

(d) $E_i : \mathbb{R} \times \mathscr{M}^{\infty} \to \mathbb{R}$ is *division i's evaluation measure.*

Under certain iterative procedures, the Center may be able to take a "good" decision after a finite number of iterations. Since x is defined only for an infinite string of messages, it would appear that the control mechanism cannot handle situations in which the Center has collected sufficient information for optimal decision making after a finite number of iterations of the communication process. That is, if (1a)–(1c) are viewed as partially defining a (decentralized) nonlinear programming algorithm, what happens if the algorithm converges to a solution of (A) in a finite number of iterations? For the procedures considered here it will suffice to assume that the Center acts as though the final message of each division has been repeated an infinite number of times. Formally,

$$x: \mathscr{M}^t \times (m_t \times m_t \times \cdots) \to X. \tag{2}$$

A stopping criterion is associated with any iterative procedure: if this criterion is satisfied at iteration t, the Center's decision is $x(m^t, (m_t \times m_t \times \cdots))$.

[9] $\mathscr{M}^{\infty}$ is the Cartesian product of $\mathscr{M}$ with itself an infinite number of times.

It would then remain to be shown for a particular procedure that, if the stopping criterion is satisfied at iteration t, the decision $x(m^t, (m_t \times m_t \times \cdots))$ is optimal.

IV. The Division's Problem

Once the control mechanism C has been specified, the problem for each division is to decide how to choose the messages $m_{it} \in \mathcal{M}$ to be sent to the Center at each iteration. Each division wants to send messages to the Center that induce the Center to take a decision that maximizes the division's evaluation. Thus it chooses, for each iteration t, a *response rule*, or a function u_{it}, that determines at iteration t how it should respond to the messages $(y^1,..., y^t)$ from the Center (and implicitly to the past messages of the other divisions, since each y^t, for $t \geqslant 2$, depends on the messages of all divisions for all previous iterations).

$$u_{it}\colon \bigtimes_{\tau=1}^{t} Y \to \mathcal{M}. \tag{3}$$

A *strategy* u_i is an infinite sequence of response rules, one for each possible iteration of the communication process:

$$u_i \equiv \{u_{it}\}_{t=1}^{\infty}. \tag{4}$$

Each message of any division can be expressed in terms of the messages (and thus the response rules) of all the divisions at previous iterations, the response rule of the division at the current iteration, the messages of the Center at previous iterations, and the starting point y^1. Since associated with each control mechanism is a unique starting point y^1 and a set of message rules $y^t(\cdot)$ for $t \geqslant 2$, a control mechanism C and the N-tuple of divisional strategies $u = (u_1,..., u_N)$ uniquely determine the infinite sequence of messages that the Center will receive. The available strategies for a division to choose among are members of the set

$$U \equiv \{\text{the space of infinite sequences defined by (3) and (4)}\}. \tag{5}$$

Given the joint strategy $u = (u_1,..., u_N)$, the infinite string of messages can be defined recursively as follows (where the dependence on the control mechanism C is suppressed):

$$m_{it} : U^N \to \mathcal{M}, \tag{6}$$

where

$$\begin{aligned} m_{i1}(u) &\equiv u_{i1}(y^1), \\ m_{i2}(u) &\equiv u_{i2}(y^1, y^2(m^1(u))), \\ &\vdots \\ m_{it}(u) &\equiv u_{it}(y^1, y^2(m^1(u)),..., y^t(m^{t-1}(u))), \end{aligned}$$

and

$$m^{\infty}: U^N \to \mathscr{M}^{\infty},$$

where

$$m^{\infty}(u) \equiv (m_{11}(u), \ldots, m_{N1}(u), m_{12}(u), \ldots, m_{N2}(u), \ldots).$$

The payoff for division i, as a function of the vector of strategies u, is division i's evaluation when the messages are $m^{\infty}(u)$:[10]

$$W_i[u] \equiv E_i(\Pi_i(x(m^{\infty}(u))), m^{\infty}(u)). \tag{7}$$

V. Definition of Optimality

The choice of divisional strategies can be thought of in game-theoretic terms. The payoff functions $W_i[\cdot]$ and the set of divisional strategies U define an N-person (non-zero sum) game in normal form, which we assume is played noncooperatively.

A vector of strategies u^* will be called a "*solution*" to the game defined by the payoff functions $W_i[\cdot]$ and the set of strategies U if

u^* is a Nash equilibrium and $W_i[u^*] \geqslant W[u]$ for any other Nash equilibrium u, for all i.[11,12] (8)

An iterative control mechanism will be called *optimal* if a "solution" u^* exists and if any "solution" leads the Center to take a decision that solves problem (A). Formally, the iterative control mechanism

$$C^* \equiv (\mathscr{M}^*, \{y^{t^*}(\cdot)\}_{t=1}^{\infty}, x^*(\cdot), \langle E_i^*(\cdot)\rangle_{i=1}^{N}) \tag{9}$$

is an optimal control mechanism if the following two properties are satisfied:

Decisiveness: There exists a u^* satisfying (8).

Efficiency: If u^* satisfies (8) then $x^*(m^{\infty}(u^*))$ solves problem (A).

At first glance problem (A) might seem limited in its applicability, since the only decisions are made by the Center. However, it is easily extended to

[10] Since each u generates a sequence of messages $m^{\infty}(u)$, $x(\cdot)$ can be calculated for every u, since $x: \mathscr{M}^{\infty} \to X$.

[11] $\tilde{u}$ is a Nash equilibrium iff $W_i[\tilde{u}] \geqslant W_i[\tilde{u}/u_i]$ for all $u_i \in U$, for all i.

[12] Since, in general, intermixtures $(\hat{u}_1, u_2^*, \ldots, \hat{u}_i, \ldots, u_j^*, \ldots, \hat{u}_N)$ of Nash equilibria $\hat{u}$ and u^* are *not* Nash equilibria, to avoid ambiguity the divisions should be able to choose among multiple Nash points. The solution concept for this game is thus strengthened by requiring the existence of a best Nash equilibrium.

cover models in which both the Center and the divisions must take decisions, when it is assumed that the Center takes its decision first and that the divisions take their decisions on the basis of the Center's decision.[13] (For a complete demonstration of this fact, see Cohen [3].)

VI. Definition of a Class of Optimal Control Mechanisms—$\mathscr{C}^*$

In this section a special class of control mechanisms is examined. The specification of this class requires a few preliminary definitions:

$$\mathscr{H} \equiv \{H_i \mid H_i : \mathbb{R}^L \to \mathbb{R}\},$$
$$\mathscr{H}_0 \equiv \{H_i \mid H_i \in \mathscr{H} \text{ and } H_i \text{ is concave}\}.$$

A strategy u_i was defined as an infinite sequence of functions u_{it}, where $u_{it} : \mathsf{X}_{\tau=1}^{t} Y \to \mathscr{M}$. A particular type of strategy is one in which $u_{it} : Y \to \mathscr{M}$ for all t and u_{it} is identical for all t, i.e., $u_{it} = u_{is}$ for all $t, s \in [1, \infty)$. The strategy u_i is thus a stationary strategy, in the language of dynamic programming.

$$\tilde{U} \equiv \{u_i \mid u_i \in U, u_{it} : Y \to \mathscr{M} \text{ and } u_{it} = u_{is} \text{ for all } t, s \in [1, \infty)\}.$$

(The tth element of the infinite sequence u_i, when $u_{it} = u_{is}$ for all $t, s \in [1, \infty)$, will be referred to as u_i. Therefore $u_i(y^t)$ should be interpreted as $u_{it}(y^t)$.)

The class of control mechanisms considered can first be characterized by their operating rules. In the previous work on incentive mechanisms, most notably that of Groves [10] and Groves and Ledyard [12], the following question is never addressed: How does the Center actually calculate the vector x once it has received the divisions' reported profit contribution functions. It is implicitly assumed that the Center can find a solution to the resultant mathematical programming problem. Under such an implicit assumption, it is immediate that the Center can find a solution to its overall

[13] If the divisions and the Center make their respective decisions simultaneously, the Center's problem will not necessarily be solved by these decisions. Suppose that the decisions were taken simultaneously. It would not necessarily be true that each division's decision and the Center's decision would be *jointly* feasible for that division. If the divisions take their decisions first there will not necessarily be *one* decision for the Center that is jointly feasible with the local decisions of all the divisions. When the Center takes its decision first, local decisions of the divisions and the Center's decision will be jointly feasible for all divisions. Sequencing of decision making allows sufficient information to be exchanged so that the Center's problem can be solved. Groves [9] deals with the problem of simultaneous decision making by the Center and the divisions as a problem in decision making under uncertainty.

problem if the divisions report their true profit contribution functions. With iterative communication, the calculation of the vector x is an inherent part of the control problem. Restriction 1 merely ensures the existence of divisional strategies that will lead the Center to calculate the solution to problem (A).

Restriction 1. Given the iterative control mechanism C, there exists a non-empty-valued mapping

$$\Psi: \mathscr{H} \to \tilde{U}$$

such that if $H_i \in \mathscr{H}_0 \subset \mathscr{H}$ for all i, then for any $u_i \in \Psi[H_i]$,

$$x(m^\infty(u)) \text{ maximizes } \sum_{i=1}^{N} H_i(x) + \Pi_0(x) \text{ subject to } x \in X.$$

A *truthful strategy for i* will be any $u_i \in \Psi[\Pi_i]$.

In order to prove the dominant strategy property of "telling the truth" (i.e., reporting true profit contribution functions) the work on noniterative procedures requires the Center to be able to calculate the value of reported profit at the Center's decision x. This calculation is straightforward since the Center knows the *entire* reported profit contribution function of each division. When communication is iterative, the Center can only approximate this function. Restriction 2 ensures that when the decision $x(\cdot)$ is taken, the Center has an exact approximation of each division's reported profit contribution at $x(\cdot)$. The rules defined below in Restriction 2 are approximation rules for the reported profit contribution functions of the divisions.

Restriction 2. Given the iterative control mechanism C, there exist rules $P_i : X \times \mathscr{M}^\infty \to \mathbb{R}$ such that

If $u_i \in \Psi[H_i]$, where $H_i \in \mathscr{H}_0$, then for all $u \backslash u_i \in U^{(N-1)}$, $P_i(x(m^\infty(u)), m_i^\infty(u)) = H_i(x(m^\infty(u)))$, where $m_{it}(u) = u_i(y^t(m^{t-1}(u)))$ for all t, and all i.[14]

It cannot be *assumed* that a division will play a particular type of strategy; it must be shown that playing this type of strategy is in the best interests of the division. Evaluation measures are used to provide the divisions with an incentive to play a particular type of strategy. What can be required is that the messages sent to the Center by the divisions obey some consistency requirements.

Given an iterative control mechanism C satisfying Restriction 1, the infinite sequence of messages m_i^∞ (10)

[14] $(u\backslash u_i) \equiv (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_N)$.

is said to be *consistent with* $H_i(\cdot)$ *given* $\{y^t\}_{t=1}^{\infty}$ if there exists a $u_i \in U$ such that

(a) $u_i \in \Psi[H_i]$ and

(b) $m_{it} = u_i(y^t)$ for all t.

Thus, the sequence of messages m_i^{∞} is *consistent with a concave function given* $\{y^t\}_{t=1}^{\infty}$ if there exist a concave function H_i and a strategy u_i satisfying (10).[15,16]

Restriction 3. Given the iterative control mechanism C satisfying Restrictions 1 and 2, $E_i(\Pi_i, m^{\infty}) \equiv F_i(R_i(\Pi_i, m^{\infty}))$, where

$$R_i(\Pi_i, m^{\infty}) \equiv \begin{cases} \Pi_i + \sum_{j \neq i} P_j(x(m^{\infty}), m_j^{\infty}) + \Pi_0(x(m^{\infty})) \\ \qquad \text{if } m_i^{\infty} \text{ is consistent with a concave function} \\ \qquad \text{given } \{y^t\}_{t=1}^{\infty}\text{; } and\ \Pi_i \geqslant P_i(x(m^{\infty}), m_i^{\infty}) \\ -\infty \qquad \text{otherwise} \end{cases}$$

and $F_i : \mathbb{R} \to \mathbb{R}$ predetermined and strictly increasing.[17]

Since $E_i(\cdot)$ must be defined for all m^{∞}, whether or not the rules of the game have been followed, the $P_j(\cdot)$ functions must be defined for *all* m_j^{∞}. The inclusion of the condition that actual profit contributions be at least as large as reported profit contribution is not as innocuous as it may seem at first glance. This condition creates a discontinuity in the evaluation measure, since any deviation (even a small one) downward from reported profits will cause the division to have an evaluation of $-\infty$. This definition also requires ex post observation by the Center of divisional contribution. Thus, given Restriction 2, these evaluations are a type of "profit sharing." Why not just define $E_i(\cdot)$ to be some fixed fraction of actual profit, i.e., define $E_i(\Pi_i, m^{\infty}) \equiv \alpha_i(\sum \Pi_i(x(m^{\infty})))$, $0 < \alpha_i < 1$? Evaluations of the type defined by Restriction 3 are not dependent upon the local decisions of the other divisions. Thus the optimal strategy for division i does not depend on the rationality of the other divisions. Profit sharing does not have the independence property. Since this model is easily generalizable to situations that include divisional decision making, profit sharing is not a viable alternative.

[15] In general, given $\{y^t\}_{t=1}^{\infty}$, m_i^{∞} may be consistent with many different H_i, some in $\mathscr{H}_0$ and some not. It is thus not possible in general to determine if i is playing according to some $H_i \in \mathscr{H}_0$ by observing only m_i^{∞} and $\{y^t\}_{t=1}^{\infty}$.

[16] If m_i^{∞} is consistent with a function $H_i \in \mathscr{H}_0$ for *all* sequences $\{y^t\}_{t=1}^{\infty}$, where $y^t \in Y$, then division i *must* be playing according to the strategy $u_i \in \Psi[H_i]$, where $m_{it}(y^t) = u_i(y^t)$.

[17] It may not be decidable in general whether or not a particular infinite string of messages m_i^{∞} is consistent with some $H_i(\cdot)$ given $\{y^t\}_{t=1}^{\infty}$. Thus, Restriction 3 implicitly excludes mechanisms for which this is not decidable.

The class of iterative control mechanisms defined by Restrictions 1–3 will be denoted by $\mathscr{C}^*$:

$$\mathscr{C}^* \equiv \{C \mid C \text{ is an iterative control mechanism and } C \text{ satisfies Restrictions 1–3}\}.$$

VII. Properties of the Class $\mathscr{C}^*$

An optimal control mechanism has been defined as a control mechanism for which:

(a) there exists a "best" Nash equilibrium that dominates (in terms of divisional payoff) all other Nash equilibria for every division; and

(b) any "best" Nash equilibrium leads to an optimal decision by the Center.

The class of mechanisms $\mathscr{C}^*$ can be shown to be optimal. It is not possible to show the existence of a dominant strategy equilibrium. However, if all players are restricted, a priori, to playing strategies associated with concave functions H_i, i.e., $u_i \in \Psi[\mathscr{H}_0] \equiv \tilde{U}^*$, not only does a best Nash equilibrium exist, but this best Nash equilibrium is also a dominant strategy equilibrium for the restricted class.

We define the equivalence class of truthful strategies for division i as the set of strategies $\mathscr{U}_i^*$, where

$$\mathscr{U}_i^* \equiv \Psi[\Pi_i + c_i] \qquad \text{for all constants } c_i. \tag{11}$$

A restricted equivalence class of truthful strategies for division i is defined by restricting $\Pi_i(\cdot)$ to the domain X. If $\bar{\Pi}_i$ is the restricted form of Π_i, then

$$\bar{\mathscr{U}}_i^* \equiv \Psi[\bar{\Pi}_i + c_i] \qquad \text{for all constants } c_i.$$

The following lemma is immediate from Restrictions 1 and 3 and (11), and is given without proof. Unless otherwise stated, the proofs of all subsequent lemmas and theorems appear in the Appendix.

Lemma 1. *Given any iterative control mechanism C satisfying Restrictions* 1 *and* 3, *if*

$$u_i^1, u_i^2 \in \mathscr{U}_i^*$$

then

$$W_i[u/u_i^1] = W_i[u/u_i^2] \qquad \textit{for all } u\backslash u_i \in U^{(N-1)}, \qquad \textit{for all } i.^{18}$$

[18] $u/\tilde{u}_i \equiv (u_1, \ldots, u_{i-1}, \tilde{u}_i, u_{i+1}, \ldots, u_N)$.

THEOREM 1. *Given a control mechanism* $C \in \mathscr{C}^*$, *for all* $u \backslash u_i \in \tilde{U}^{*(N-1)}$

$$W_i[u/u_i^*] \geqslant W_i[u] \qquad \textit{for all} \quad u_i \in \tilde{U}^* \text{ iff } u_i^* \in \mathscr{U}_i^*.$$

Thus, a dominant strategy equilibrium exists when the divisions are restricted to playing strategies in $\tilde{U}^*$. Does there exist a strategy in $\tilde{U}^*$ that is best against a particular $(N-1)$-tuple of strategies that is in $U^{(N-1)}$ but not in $\tilde{U}^{*(N-1)}$? If division i wishes to maximize its payoff, it must send messages m_i^∞ to the Center that are consistent with a concave function given $\{y^t\}_{t=1}^\infty$; this ensures the existence of a best strategy in $\tilde{U}^*$ for division i. The following is straightforward, and is given without proof:

LEMMA 2. *Given a control mechanism* $C \in \mathscr{C}^*$, *for each* $u \backslash u_i \in U^{(N-1)}$ *there exists a particular* $\hat{u}_i \in \tilde{U}^*$ *such that*

$$W_i[u/\hat{u}_i] \geqslant W_i[u] \qquad \forall u_i \in U. \tag{12}$$

The proof of the (if) part of Theorem 1, i.e., the proof that for all $u \backslash u_i \in \tilde{U}^{*(N-1)}$, $W_i[u/u_i^*] \geqslant W_i[u]$ for all $u_i \in \tilde{U}^*$, relies on the following two conditions:

(a) $x(m^\infty(u/u_i^*))$ maximizes

$$\Pi_i(x) + \sum_{j \neq i} H_j(x) + \Pi_0(x) \text{ over } X; \tag{13}$$

(b) $P_j(x(m^\infty(u/u_i^*)), m_j^\infty(u/u_i^*)) = H_j(x(m^\infty(u/u_i^*)))\ \forall j.$

If $u \backslash u_i \notin \tilde{U}^{*(N-1)}$ the conditions in (13) do not *necessarily* hold. Therefore it is not possible to prove that for any $u \backslash u_i \in U^{(N-1)}$ condition (12) holds.

Theorem 1 establishes the fact that the "truth" is a Nash equilibrium, and in addition that the "truth" is the best concave Nash equilibrium. Lemma 1 establishes that there exists, for each particular $(N-1)$-tuple of divisional strategies, a best concave strategy ? (This is the result of the divisions sending messages consistent with a concave strategy.)

Is there reason to believe that the divisions will play the particular Nash equilibrium of "truthful" strategies? The answer is in the affiffirmative for the following reasons: first, the divisions already know their respective "truthful" strategies, and playing one of those strategies requires no additional calculation on their part. Second, to play any other Nash equilibrium vector of strategies, each division must know what the other divisions are doing, in order to calculate the best strategy in the particular situation. Thus, the N-tuple of "truthful" strategies is the only Nash equilibrium that is "independent" of the other divisions' strategies.

LEMMA 3. *For any control mechanism* $C \in \mathscr{C}^*$ *any* N-*tuple of strategies* u^*, *where* $u_i^* \in \mathscr{U}_i^*\ \forall i$, *is a Nash equilibrium.*

LEMMA 4. *For any control mechanism $C \in \mathscr{C}^*$ any Nash equilibrium u^*, where $u_i^* \in \mathscr{U}_i^*$ $\forall i$, is a "solution," i.e.,*

$$W_i[u^*] \geqslant W_i[u] \qquad \forall \text{ Nash equilibria } u, \forall i.$$

LEMMA 5. *For any control mechanism $C \in \mathscr{C}^*$, if $\bar{u}$ is a "solution," then $x(m^\infty(\bar{u}))$ solves problem* (A).

The major result of this section is stated in the following theorem:

THEOREM 2. *Any control mechanism C in the class $\mathscr{C}^*$ is* optimal.

In the previous work that has been done on incentive mechanisms no specific assumptions have been made about the Center's ex post knowledge of actual divisional profit contribution. The divisions are assumed to send entire profit contribution functions, and it can be shown that the N-tuple of "truthful" strategies is a dominant strategy equilibrium. As was pointed out above, it is assumed in the previous work that the Center possesses some method of solving the mathematical programming problem it receives from the divisions. The important point is that the Center knows each division's *entire* strategy, and thus can calculate a global solution to the reported problem. When communication is iterative in nature, the Center knows only a local approximation of the strategies near the decision x, and is not guaranteed a global solution. Thus, if some of the divisions play strategies that do not ensure convergence to a solution (i.e., nonconcave strategies), the truth may not be the best strategy for the remaining divisions. The lack of complete information about strategies ex post leads to a weaker result than can be obtained about the "complete information," or noniterative, mechanisms.

Finally, we must give an interpretation of the evaluation measures $E_i(\cdot)$. One alternative would be to view the functions $R_i(\cdot)$ as providing an index number and the transformation $F_i(\cdot)$ as translating that number into the argument of the division manager's utility function. For example, the manager of division i may receive satisfaction from paid vacation days, while the manager of division j wants to maximize the probability of receiving a key to the executive lavatory.

Alternatively, the evaluation measures and thus the payoff functions, may represent real resource flows. In that case, one must worry about the question of a "balanced budget" at any equlilibrium.

LEMMA 6. *If $F_i(\cdot) \equiv \alpha_i$ for all i, where $0 < \alpha_i < 1$ and $\sum_{i=1}^N \alpha_i = 1$, then, at any Nash equilibrium u,*

$$\sum_{i=0}^{N} \Pi_i(x(m^\infty(u))) \geqslant \sum_{i=1}^{N} W_i[u].$$

VIII. Applications

One may rightfully ask the following question: Is the set $\mathscr{C}^*$ of optimal control mechanisms nonempty? If so, how exclusionary are Restrictions 1 and 2 on the class of admissible operating rules?

The answer to the first question is yes; it is straightforward to verify that the Dantzig–Wolfe decomposition algorithm [4], as well as Malinvaud's "Procedure Implying Mathematical Programming at the Center" [18] satisfy Restrictions 1 and 2 (see Cohen [3]). It can also be shown that two simple nonlinear algorithms satisfy the restrictions: an outer-linearization algorithm and an inner-linearization (price decomposition) algorithm (see Cohen [3]). (Both the Dantzig–Wolfe algorithm and Malinvaud's procedure are special cases of the latter.)

Restriction 1 requires only that the operating rules actually define a convergent algorithm for solving a nonlinear concave programming problem. Restriction 2 limits attention to procedures that calculate the value of the objective function at the solution. The inner- and outer-linearization algorithms cited above satisfy Restriction 2, thus leading one to postulate that a wide range of algorithms satisfy the restriction.

Appendix

For ease of exposition, we make the following definitions:

$$x(m^\infty(u)) \equiv x(u),$$

$$A_i(x) \equiv F_i\Big(\Pi_i(x) + \sum_{j\neq i} P_j(x, m_j^\infty) + \Pi_0(x)\Big),$$

$$B_i(x) \equiv F_i\Big(\Pi_i(x) + \sum_{j\neq i} H_j(x) + \Pi_0(x)\Big),$$

$$C_i(x) \equiv F_i\Big(\Pi_i(x) + \sum_{j\neq i} \Pi_j(x) + \Pi_0(x)\Big).$$

Then by definition

$$W_i[u] = A_i(x(u)).$$

(Bars over $A_i(\cdot)$, $B_i(\cdot)$, and $C_i(\cdot)$ denote a bar over $\Pi_i(\cdot)$).

Proof of Theorem 1. (If) Let $u_i^* \in \Psi[\bar{\Pi}_i]$. From Lemma 1, $W_i[u/u_i'] = W_i[u/u_i^*]$ for all $u_i' \in \mathscr{U}_i^*$. We therefore need only prove the theorem for u_i^*. From Restriction 2,

$$A_i(x(u/u_i^*)) = B_i(x(u/u_i^*)).$$

From Restriction 1 and $F_i(\cdot)$ strictly increasing,

$$\bar{B}_i(x(u/u_i^*)) \geqslant \bar{B}_i(x) \qquad \forall x \in X.$$

Since $\Pi_i(x) = \bar{\Pi}_i(x)\ \forall x \in X$,

$$B_i(x(u/u_i^*)) = \bar{B}_i(x(u/u_i^*)) \geqslant \bar{B}_i(x) = B_i(x) \qquad \forall x \in X.$$

In particular,

$$B_i(x(u/u_i^*)) \geqslant B_i(x(u)) \qquad \forall u_i \in U.$$

Since $u_j \in \tilde{U}^*$ for all $j \neq i$, Restriction 2 gives

$$B_i(x(u)) = A_i(x(u)) \qquad \forall u_i \in U.$$

Therefore for all $u \backslash u_i \in \tilde{U}^{*(N-1)}$

$$W_i[u/u_i^*] \geqslant W_i[u] \qquad \forall u_i \in U.$$

(Only if) Let $\tilde{u}_i \in \tilde{U}^*$ be such that for all $u \backslash u_i \in \tilde{U}^{*(N-1)}$

$$W_i[u/\tilde{u}_i] = W_i[u/u_i^*] \geqslant W_i[u] \qquad \forall u_i \in U$$

(where $u_i^* \in \Psi[\Pi_i]$).

We must show that $\tilde{u}_i \in \mathscr{U}_i^*$. From the (if) part of this theorem, its proof, and Restriction 2 we get

$$W_i[u/\tilde{u}_i] = B_i(x(u/\tilde{u}_i)) \geqslant B_i(x) \qquad \forall x \in X.$$

Restriction 1 yields $x(m^\infty(u/\tilde{u}_i))$ as the maximizer of

$$\tilde{H}_i(x) + \sum_{j \neq i} H_j(x) + \Pi_0(x) \qquad \text{s.t.} \quad x \in X,$$

For the assertion of the theorem to be true for all $u \backslash u_i$ associated with concave functions, it must be true for all $u \backslash u_i$ associated with strictly concave, differentiable functions. If $D_{\Pi_i} = \{x \mid \Pi_i(x) \text{ is differentiable}\}$ and $D_{\tilde{H}_i}$ and D_{Π_0} are defined similarly, then concavity of Π_i, $\tilde{H}_i$, and Π_0 implies that

(a) $D_{\Pi_i}^C$, $D_{\tilde{H}_i}^C$, and $D_{\Pi_0}^C$ (where C denotes complement) are all sets of measure zero; and

(b) D_{Π_i}, $D_{\tilde{H}_i}$, and D_{Π_0} are each contained in the interior domain of their respective functions of definition. (See Rockafellar [21, p. 246, Theorem 25.5].)

Therefore,

$$(D_{\Pi_i} \cap D_{\tilde{H}_i} \cap D_{\Pi_0})^C = (D_{\Pi_i}^C \cup D_{\tilde{H}_i}^C \cup D_{\Pi_0}^C)$$

is a set of measure zero. For each $\tilde{x} \in D_{\Pi_i} \cap D_{\tilde{\Pi}_i} \cap D_{\Pi_0} \cap X$ define the strategies $\tilde{u} \backslash u_i$ such that $\tilde{u}_j \in \Psi[\tilde{\Pi}_j]$ and

$$\sum_{j \neq i} \tilde{\Pi}_j(\cdot) \text{ is strictly concave and differentiable;}$$

$$\nabla \sum_{j \neq i} \tilde{\Pi}_j(\tilde{x}) = -\nabla \Pi_i(\tilde{x}) - \nabla \Pi_0(\tilde{x}). \tag{14}$$

(To show the existence of such strategies we need only show the existence of the $\tilde{\Pi}_j(\cdot)$, which is obvious.) A necessary and sufficient condition for a unique interior maximum of a strictly concave, differentiable function $f(\cdot)$ at x^* is that $\nabla f(x^*) = \mathbf{0}$. Since

$$\nabla \Pi_i(\tilde{x}) + \nabla \sum_{j \neq i} \tilde{\Pi}_j(\tilde{x}) + \nabla \Pi_0(\tilde{x}) = \mathbf{0}$$

$\tilde{x}$ is the unique maximizer of $\Pi_i(x) + \sum_{j \neq i} \tilde{\Pi}_j(x) + \Pi_0(x)$ over X. In addition, $\tilde{x} \in$ {interior domain of $\tilde{\Pi}_i(\cdot)$} and $\tilde{x}$ maximizes $\sum_{i=1}^{N} \tilde{\Pi}_i(x) + \Pi_0(x)$ over X. Therefore,

$$\nabla \sum_{i=1}^{N} \tilde{\Pi}_i(\tilde{x}) + \nabla \Pi_0(\tilde{x}) = \mathbf{0}.$$

This gives the result that

$$\nabla \tilde{\Pi}_i(\tilde{x}) = \nabla \Pi_i(\tilde{x}).$$

Since functions $\tilde{\Pi}_j(\cdot)$, $j \neq i$, satisfying (14) can be defined for each $x \in D_{\Pi_i} \cap D_{\tilde{\Pi}_i} \cap D_{\Pi_0} \cap X$,

$$\nabla \tilde{\Pi}_i(\tilde{x}) = \nabla \Pi_i(\tilde{x}) \qquad \text{a.e. in } X.$$

Since $\Pi_i(\cdot)$ and $\tilde{\Pi}_i(\cdot)$ are concave and defined over all of X, they are necessarily continuous over X (see Rockafellar [21, p. 83, Corollary 10.11]). Therefore,

$$\tilde{\Pi}_i(x) = \Pi_i(x) + c_i \qquad \text{for all } x \in X, \text{ where } c_i \text{ is a constant.}$$

This gives the result that

$$\tilde{u}_i \in \mathscr{U}_i^*. \quad \blacksquare$$

Proof of Lemma 3. From the (if) part of the proof of Theorem 1 and Restriction 2, for all $u_i \in U$

$$W_i[u^*] \geqslant C_i(x(u^*/u_i)) = A_i(x(u^*/u_i)) = W_i[u^*/u_i]. \quad \blacksquare$$

Proof of Lemma 4. Let $\tilde{u}$ be any Nash equilibrium. Then for any i

(1) $m_i^\infty(\tilde{u})$ is consistent with a concave function given $\{y^t\}_{t=1}^\infty$; and

(2) $\Pi_i \geqslant P_i(x(\tilde{u}), m_i^\infty(\tilde{u}))$.

Thus, for all i, if $u_i^* \in \Psi[\Pi_i]$ for all i,

$$W_i[u^*] = C_i(x(u^*)) \geqslant C_i(x(\tilde{u})) \geqslant A_i(x(\tilde{u})) = W_i[\tilde{u}]. \quad \blacksquare$$

Proof of Lemma 5. Since any u^*, such that $u_i^* \in \mathscr{U}_i^* \, \forall i$, is a solution,

$$W_i[\tilde{u}] = W_i[u^*] \qquad \forall i.$$

Assume $x(\tilde{u})$ does not solve problem (A), i.e., that

$$\sum_{i=0}^{N} \Pi_i(x(u^*)) > \sum_{i=0}^{N} \Pi_i(x(\tilde{u})),$$

where $u_i^* \in \Psi[\Pi_i]$.

From the proof of Theorem 1, $F_i(\cdot)$ strictly increasing, Restriction 1, and $\tilde{u}$ a Nash equilibrium (and thus $\Pi_i \geqslant P_i(x(\tilde{u}), m_i^\infty(\tilde{u}))$ for all i) we get

$$W_i[\tilde{u}] = W_i[u^*] = C_i(x(u^*)) > C_i(x(\tilde{u})) \geqslant A_i(x(\tilde{u})) = W_i[\tilde{u}],$$

which is a contradiction.

Therefore, $x(\tilde{u})$ must solve problem (A). $\blacksquare$

Proof of Lemma 6. Let u be a Nash equilibrium. Then

$$\Pi_i(x(u)) \geqslant P_i(x(u), m_i^\infty(u)) \qquad \forall i.$$

Therefore

$$\begin{aligned}
\sum_{i=1}^{N} W_i[u] &= \sum_{i=1}^{N} \alpha_i \left[\Pi_i(x(u)) + \sum_{j \neq i} P_j(x(u), m_j^\infty(u)) + \Pi_0(x(u))\right] \\
&\leqslant \sum_{i=1}^{N} \alpha_i \left[\sum_{j=0}^{N} \Pi_j(x(u))\right] \\
&= \sum_{i=0}^{N} \Pi_i(x(u)). \quad \blacksquare
\end{aligned}$$

References

1. K. Arrow, Control in large organizations, *Management Sci.* **10**, No. 3 (1964).
2. K. Arrow and L. Hurwicz, Decentralization and computation in resource allocation, *in* "Essays in Economics and Econometrics" (R. Pfouts, Ed.), pp. 34–104, Univ. North Carolina Press, Chapel Hill, 1960.
3. S. Cohen, "Incentive Compatible Control of the Multidivisional Firm with Iterative Communication," unpublished Ph.D. thesis, Northwestern University, Evanston, Ill., 1977.

4. G. Dantzig and P. Wolfe, The decomposition algorithm for linear programs, *Econometrica* **29**, No. 4 (1961), 767–778.
5. J. Dreze and D. de la Vallee Poussin, A Tâtonnement process for public goods, *Rev. Econ. Stud.* **38**, No. 2 (1971), 133–150.
6. A. Geoffrion, Elements of large scale mathematical programming: Parts I and II, *Management Sci.* **16**, No. 11 (1970), 652–701.
7. R. Goldberg, "Methods of Real Analysis," Chap. 7, Xerox College Publishing, Lexington, Mass., 1964.
8. J. Green and J. J. Laffont, Characterization of satisfactory mechanisms for the revelation of preferences for public goods, *Econometrica*, **45**, No. 2 (1977), 427–438.
9. T. Groves, Incentives in teams, *Econometrica* **41**, No. 4 (1973), 617–631.
10. T. Groves, "Incentive Compatible Control of Decentralized Organizations," Working Paper No. 166, Center for Mathematical Studies in Economics and Management Science, Northwestern University, Evanston, Ill., August 1975.
11. T. Groves, Information, incentives, and the internalization of production externalities, *in* "Theory and Management of Economic Externalities" (S. Lin, Ed.), Academic Press, New York, 1975.
12. T. Groves and J. Ledyard, Optimal allocation of public goods: A solution to the "Free Rider" Problem, *Econometrica* **45**, No. 4 (1977), 783–810.
13. T. Groves and M. Loeb, Incentives and public inputs, *J. Public Econ.* **4**, No. 3 (1975), 211–226.
14. G. M. Heal, "The Theory of Economic Planning," North-Holland, Amsterdam, 1973.
15. L. Hurwicz, Optimality and informational efficiency in resource allocation processes, *in* "Mathematical Models in the Social Sciences" (K. Arrow, S. Karlin, and P. Suppes, Eds.), pp. 27–46, Stanford Univ. Press, Stanford, Calif., 1960.
16. M. Loeb, "Coordination and Information Incentive Problems in the Multi-Divisional Firm," unpublished Ph.D. thesis, Northwestern University, Evanston, Ill., 1975.
17. R. D. Luce and H. Raiffa, "Games and Decisions," Wiley, New York, 1957.
18. E. Malinvaud, Decentralized procedures for planning, *in* "Activity Analysis in the Theory of Growth and Planning" (E. Malinvaud and M. O. L. Bacharach, Eds.), pp. 170–208, Macmillan, London, 1967.
19. J. Marschak and R. Radner, "Economic Theory of Teams," Yale Univ. Press, New Haven, Conn., 1970.
20. G. Owen, "Game Theory," Saunders, Philadelphia, 1968.
21. R. T. Rockafellar, "Convex Analysis," Princeton Univ. Press, Princeton, N. J., 1970.

Printed by the St. Catherine Press Ltd., Tempelhof 37, Bruges, Belgium

Notes* II

MANAGEMENT SCIENCE
Vol. 26, No. 9, September 1980
Printed in U.S.A.

SIMPLE INEQUALITIES FOR MULTISERVER QUEUFS†

MATTHEW J. SOBEL‡

Simple inequalities are obtained for some operating characteristics of multiserver queueing models. "Loss system" and "delay system" results are presented.
(QUEUES; MULTISERVER; INEQUALITIES)

1. Introduction

It is difficult to obtain explicit formulae for operating characteristics of queueing models with more than one server. When formulae can be obtained, often they are complicated and depend on particular probability distributions. The results below are nonparametric and simple in form. Simplicity typically implies that an inequality is not sharp (cf. Kingman [6]). However, one of the principal inequalities below is simple *and* sharp. See Brumelle [2], [3] and his references for other inequalities for multiserver queueing models.

In a "loss system" with a Poisson input process, i.e., an $M/G/c/N$ model (so arriving customers are refused entry when there are N customers already inside the facility), let ρ denote the "traffic intensity" and let B denote the long-run probability that the facility is full. A principal result below is

$$(1 - 1/\rho)^+ \leqslant B \leqslant 1 - 1/(\rho + 1)$$

for all c and N. For $c \geqslant 2$ and $\rho \geqslant 1.5$, $1 - 1/\rho$ seems very close to B.

* All Notes are refereed.

† Accepted by Marcel F. Neuts; received October 29, 1979. This paper has been with the author 3 weeks for 1 revision.

‡ Georgia Institute of Technology.

0025-1909/80/2609/0951$01.25

In a stable "delay system" with a renewal input process, i.e., $GI/G/c$ with $\rho < 1$, let D denote the long-run probability that all servers are busy. Then the other principal result is

$$0 \leqslant \rho - D \leqslant (1-\rho)(c-1)$$

so ρ is a good approximation to D if ρ is close to unity, i.e. "heavy traffic."

For either $GI/G/c$ or $GI/G/c/N$, let λ^{-1} be the mean inter-arrival time, μ^{-1} be the mean service time, and $\rho = \lambda/(c\mu)$. Let $p_n(t)$ be the probability that n customers are in the facility at epoch t and suppose that the limit

$$p_n = \lim_{t \to \infty} p_n(t)$$

exists for all n and that

$$\sum_{n=0}^{\infty} p_n = 1.$$

This precludes $\rho > 1$ in $GI/G/c$ but $0 < \rho < \infty$ is admissable for $GI/G/c/N$.

Let π_n^j be the probability that the jth customer to arrive finds n customers already in the system (either $GI/G/c$ or $GI/G/c/N$) and suppose that the limit

$$\pi_n = \lim_{j \to \infty} \pi_n^j$$

exists for all n. Generally, $\pi_n \neq p_n$ but a Poisson arrival process ensures $\pi_n = p_n$ for all n (Wolff [9]). Therefore, the following interpretations of D and B are known to be correct only for $M/G/c$ and $M/G/c/N$ models, respectively:

D: the long-run probability that an arriving customer must wait in a queue;

B: the long-run probability that an arriving customer is blocked from entering the facility (because it is full).

Symbols that pertain to both the loss and the delay systems are listed below:

$$D = \sum_{j \geqslant c} p_j,$$

$$S = \sum_{j=0}^{c-1} j p_j + c \sum_{j \geqslant c} p_j,$$

$$L = \sum_{j \geqslant 0} j p_j,$$

$$L_q = \sum_{j \geqslant c} (j-c) p_j,$$

$$g = \sum_{j=0}^{c-1} j p_j,$$

$$a = \lambda/\mu = c\rho.$$

These symbols have the following interpretations: L is the long-run average number in the system, L_q is the long-run average number in the queue, S is the long-run average number of busy servers, and D is the long-run probability that all the servers are busy.

2. Delay System

Let W^j and W_q^j denote the mean (expected value) lengths of time that the jth customer spends in the system and in the queue, respectively. Suppose that the limits

$$W = \lim_{j\to\infty} W^j \quad \text{and} \quad W_q = \lim_{j\to\infty} W_q^j$$

exist. Then $L = \lambda W$ and $L_q = \lambda W_q$ (Little [8]) are exploited in the following proof.

THEOREM 1.

$$S = \lambda/\mu = a; \tag{1}$$

$$D = (S - g)/c; \tag{2}$$

$$D = \rho - g/c; \tag{3}$$

$$0 \leqslant \rho - D \leqslant (1-\rho)(c-1). \tag{4}$$

PROOF. From $W = W_q + \mu^{-1}$,

$$L = \lambda W = \lambda\left(W_q + \mu^{-1}\right) = L_q + c\rho = a.$$

Also, $L = L_q + S$ so $S = a$ which is (1). By definition, $S = g + cD$ which yields (2). Then (1) and (2) give (3). For (4), $D \leqslant \rho$ from (3). Also from (3),

$$D = \rho - g/c = \rho - c^{-1}\sum_{j=0}^{c-1} jp_j$$

$$\geqslant \rho - \left[(c-1)/c\right]\sum_{j=1}^{c-1} p_j \geqslant \rho - (1-D)(c-1)/c$$

which gives the right inequality in (4). Q.E.D.

Theorem 1's most useful results are (1) and (4). We know that $D = \rho$ (exactly) in a single-server system. It follows from (4) that heavily-loaded systems, i.e. ρ close to unity, with few servers have values of D close to ρ.

Theorem 1 was first displayed in Sobel [9]. The results seemed to be new then but they are sufficiently simple that they may have been obtained by others. In fact, Harris [4] has since presented (1).

3. Loss System with Poisson Arrival Process

The model now is $M/G/c/N$ with $0 < \rho = \lambda/(c\mu) = a/c < \infty$. Let $B = p_N$ which is the long-run probability that the system is full. A Poisson arrival process implies $p_n = \pi_n$ for all n so $\pi_N = p_N$ and B is the long-run probability that an arrival is blocked from entering the facility because it is full. Hence, $D - B = \sum_{j=c}^{N-1} p_j$ is the probability that an arriving customer gains entry and must wait in the queue, and $1 - D = \sum_{j=0}^{c-1} p_j$ is the probability that an arriving customer gains entry and does not wait in the queue.

The effective arrival rate is $\lambda(1-B)$ because $p_N = \pi_N = B$ so the following simple modification of Theorem 1 applies to loss systems.

COROLLARY 1.

$$S = (1 - B)a \quad and \quad B = 1 - S/a; \tag{5}$$

$$D = \rho(1 - B) - g/c; \tag{6}$$

$$0 \leqslant \rho(1 - B) - D \leqslant (c - 1)(1 - \rho[1 - B]). \tag{7}$$

THEOREM 2.

$$(1 - 1/\rho)^{+} \leqslant B \leqslant 1 - 1/(\rho + 1); \tag{8}$$

$$1 - D \leqslant c/(1 + a). \tag{9}$$

PROOF. From $c \geqslant S$ and definitions of symbols,

$$c \geqslant S = g + cD \geqslant cD = c\left(B + \sum_{j=c}^{N-1} p_j\right) \geqslant cB.$$

These inequalities and (5) imply

$$1 - c/a \leqslant B \leqslant 1 - Bc/a$$

which yields (8). For (9), (5) and $B \leqslant D$ imply

$$a(1 - D) \leqslant a(1 - B) = \sum_{j=0}^{c-1} jp_j + cD \leqslant (c - 1)(1 - D) + cD = c - 1 + D$$

which yields (9). Q.E.D.

The lower bound in (8), $B \geqslant 1 - 1/\rho$, seems to provide an excellent approximation for B at larger values of ρ. The data in Table 1 compare (to two decimals) $1 - 1/\rho$ with values of B tabulated for $M/M/c/N$ by Kühn [7]. Kühn's tables show B only for $\rho \leqslant 2.0$. For those tabulated values, $0 \leqslant B - (1 - 1/\rho) \leqslant 0.03$, regardless of c, if $\rho \geqslant 1.5$ and $N - c \geqslant 2$.

TABLE 1

c	$N - c$	ρ	$1 - 1/\rho$	B
1	2	2.0	0.50	0.53
1	5	2.0	0.50	0.50
2	1	2.0	0.50	0.55
2	5	2.0	0.50	0.50
2	10	2.0	0.50	0.50
2	5	1.6	0.38	0.39
6	1	2.0	0.50	0.53
6	5	1.6	0.38	0.38
6	5	2.0	0.50	0.50
40	1	1.6	0.38	0.39
40	5	1.2	0.17	0.19
40	5	1.6	0.38	0.38
100	1	1.4	0.29	0.30
100	1	1.6	0.38	0.38

Dr. D. P. Heyman of Bell Telephone Laboratories observes that the bounds in (8) cannot be tightened. In an $M/M/1/1$ model, $B = \rho/(1+\rho)$ whereas in an $M/M/1/N$ model, $B = 1 - (1-\rho^N)/(1-\rho^{N+1})$ so, if $\rho < 1$, $B \to 1 - 1/\rho$ as $N \to \infty$. Also, Heyman shows [5] that the bound $B \geqslant 1 - 1/\rho$ (with B interpreted as the long-run probability that an arriving customer is blocked from entry because the facility is full) is valid without Theorem 2's hypothesis of a Poisson input process (so $\pi_N = p_N$) and that it is valid even for non-renewal input processes. Therefore, he has significantly broadened the applicability of the bound.

Whitt, [11] proposes a heavy traffic approximation for B in $G/G/c/c$ (i.e. $N - c = 0$ so there is no waiting room at all) and compares it with $1 - 1/\rho$ and other approximations. He focuses on the "peakedness" factor in the extent to which the arrival process is non-Poisson. His numerical results for $G/M/c/c$ suggest that $1 - 1/\rho$ is consistent with the heavy traffic approximation (and others he tested) except when the number of servers is very high, peakedness is very high, and ρ is relatively low. However, he has not computed B exactly for any of the cases mentioned. Also, see [12].

Whitt [11] also observes that Borovkov [1] presents a heavy traffic limit theorem for B. Borovkov shows that $B \to 1 - 1/p$ as $a \to \infty$ and $c \to \infty$ in such a way that ρ remains above unity.

Let $h = (D-B)/(1-B)$ which is

$$\begin{aligned} h &= P\{(\text{gain entry}) \cap (\text{wait in queue})\}/P\{\text{gain entry}\} \\ &= P\{\text{wait in queue} \mid \text{gain entry}\}. \end{aligned}$$

Thus h is the long-run conditional probability that an arrival must wait in queue if the arrival is not blocked from entry.

COROLLARY 2.

$$h \leqslant \rho; \tag{10}$$

$$D - B \leqslant \rho - B(\rho+1) \leqslant 1/\rho. \tag{11}$$

PROOF. The definition of h and (7) yield

$$h = (D-B)/(1-B) \leqslant D/(1-B) \leqslant \rho.$$

For (11), (6) and (8) yield

$$\begin{aligned} D - B &= \rho(1-B) - g/c - B \leqslant \rho - B(\rho+1) \\ &\leqslant \rho - (\rho+1)(1-1/\rho) = 1/\rho. \quad Q.E.D.^{1} \end{aligned}$$

[1] The exposition of this paper was improved by incorporating a referee's recommendations.
The author was partially supported by NSF Grant SOC78-05770.

References

1. BOROVKOV, A. A., *Stochastic Processes in Queueing Theory*, Springer-Verlag, New York and Berlin, 1976 (English translation).
2. BRUMELLE, S. L., "Some Inequalities for Parallel-Server Queues," *Operations Res.*, Vol. 19 (1971), pp. 402–413.

3. BRUMELLE, S. L., "A Generalization of $L = \lambda W$ to Moments of Queue Length and Waiting Times," *Operations Res.*, Vol. 20 (1972), pp. 1127–1136.
4. HARRIS, R., "The Expected Number of Idle Servers in a Queueing System," *Operations Res.*, Vol. 22 (1974), pp. 1258–1259.
5. HEYMAN, D. P., "Comment on a Queueing Inequality," Unpublished manuscript, 1979.
6. KINGMAN, J. F. C., "Inequalities in the Theory of Queues," *J. Roy. Statist. Soc. Ser. B.*, Vol. 32 (1970), pp. 102–110.
7. KÜHN, P., *Tables on Delay Systems (Tabellen für Wartesusteme)*, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Prof-Dr. Ing. A. Lotze, 1976.
8. LITTLE, J. C. C., "A Proof for the Queueing Formula: $L = \lambda W$," *Operations Res.*, Vol. 9 (1961), pp. 383–387.
9. SOBEL, M. J., "Simple Inequalities for Multiserver Queues," *CORE Discussion Paper* 7102, Catholic University of Louvain, Heverlee, Belgium, 1971.
10. WOLFF, R. W., "Poisson Arrivals See Time Averages," Unpublished manuscript, 1976.
11. WHITT, H. W., "A Heavy Traffic Approximation for General Loss Service Systems." Unpublished manuscript, 1979.

Added in proof.

12. FREDERICKS, A. A., "Congestion in Blocking Systems–A Simple Approximation Technique," *Bell System Tech. J.*, Vol. 59 (1980), pp. 805–827.

Reprinted from
THE AMERICAN ECONOMIC REVIEW

Incentives and the Choice of Optimal Plans

By SUSAN I. COHEN*

There is a vast literature in economics concerned with the specification and discussion of planning procedures for economies (for example, see Kenneth Arrow; Theodore Groves; Geoffrey Heal; Martin Loeb and Wesley Magat). This literature is generally concerned with the algorithmic side of planning, in that it deals with behavioral rules that if followed lead to a "social optimum" (from the planner's point of view). In general, the enforcement question is not dealt with; namely, is it in the best interests of the economic agents to follow the planning board's directives?[1]

If it is not in the economic agents' interests to follow the planning board's directives, then it is not necessarily true that a social optimum will actually be attained by these procedures. The existence of enforcement rules or evaluation measures that insure compliance has been shown in several rather general contexts (see, for example, Groves). The importance of the existence of such rules is well recognized in the planning literature:

> There is a very important aspect of any planning problem...[that] may be loosely described as the problem of ensuring that it is in the interests of the firms in the economy to behave as the centre requires them to.... Once an incentive structure has been specified for a planned economy, it is reasonable to suppose that firms will be concerned to choose that strategy that will maximise the rewards accruing to them under this.... Ideally one would like to devise for any planning process an associated incentive structure such that each firm would maximise its returns under this system by revealing correct information to the centre, and producing the optimum once located. [Heal, p. 214]

The purpose of this paper is the definition of just such a system for a particular class of decentralized planning procedures (and thus a particular model of a planned economy). As noted above, the existence of such a system has been previously demonstrated in other contexts. However, those results are not applicable to decentralized planning procedures since those systems require firms to communicate entire functions or sets as a one-time message. The properties of those procedures depend heavily on the assumption of noniterative communication. A general model with iterative communication is presented in my earlier paper. The results obtained there are weaker than those obtained when communication is noniterative. In particular, Groves demonstrates that when communication is noniterative, following the planning board's directives is a dominant strategy for every firm.[2] When communication is iterative, it is only possible to show that following the planning board's directives constitutes a Nash equilibrium that weakly dominates all other Nash equilibria for all firms.

In Section I, a model of an economy (based on a model due to Edmond Malinvaud) is presented. A class of planning procedures for the economy is defined in Section II. The enforcement rules for these procedures are defined in Section III, where it

*Assistant professor, College of Industrial Management, Georgia Institute of Technology. This research was partially supported by National Science Foundation Grant SOC 78-05770. I would like to thank an anonymous referee for very helpful suggestions. All errors, of course, remain my own.

[1] Jacques Dreze and D. de la Vallee Poussin, John Roberts, and Francoise Schoumaker consider these questions in tâtonnement procedures. Loeb and Magat consider the problem in a centrally planned economy in which the communication between the planing board and the firms is of a "one-time only" nature.

[2] The models in my earlier paper and in Groves are not stated as planning models, although they can easily be interpreted as such.

is also shown that, with these rules, the agents in the economy have an incentive to follow the planning board's directives. Finally, in Section IV, it is shown that Malinvaud's model and "Procedure Implying Mathematical Programming at the Center" are special cases of the model and planning procedures defined in Section II.

I. The Model of the Economy

Let us consider the following model[3] of an economy: there are $N+1$ agents, consisting of N firms (or ministries) and the planning board. There are M commodities produced in the economy. The net demand for each good produced is the final consumption of the good minus the net outputs of the good by all firms. A program P for the economy consists of a vector of final consumption, a vector of net outputs for all goods for all firms, and a vector of net demands for all goods.

The program P chosen by the planning bureau must be feasible from the consumers' point of view. Let us assume that the planning bureau knows, a priori, the set of feasible final consumption vectors. The program P must also be technologically feasible; thus the vector of net outputs for each firm must be an element of the firm's feasible production set. Each firm knows[4] its own production-possibilities set. These sets are not known to the planning bureau. Finally, the net demand for each good in the program chosen by the planning bureau must be feasible, that is, the net demand cannot exceed the initial amount of the good available in the economy (which is known by the planning bureau).

Social preferences among alternative vectors of final consumption are represented by a utility function, known by the planning bureau. A program P^1 is preferred to P^2 if the utility of final consumption in P^1 is higher than the utility of final consumption in P^2. The planning bureau seeks a feasible program in which the utility of final consumption is greater than or equal to the utility of final consumption in all other feasible programs.

Formally, the model of the economy is as follows:

(a) x_k is the final consumption of good k, $k=1,\dots,M$. Let $x\equiv(x_1,\dots,x_M)$.

(b) y_{ki} is the net output of good k by firm i, $k=1,\dots,M$, $i=1,\dots,N$.

$y^k\equiv\sum_{i=1}^{N} y_{ki}$ is the total output of good k.

$y_i\equiv(y_{1i},\dots,y_{Mi})$ is the vector of net outputs of all goods by firm i. Let $y\equiv(y_1,\dots,y_N)$.

(c) Y_i is firm i's feasible production set, $i=1,\dots,N$. W_k is the initial amount of good k available in the economy, $k=1,\dots,M$.

(d) X is the set of feasible final consumption vectors.

(e) $U(x)$ is the utility of the final consumption x. $U(\cdot)$ is a function from the commodity space $\mathbb{R}^M$ (the M-fold Cartesian product of the real line) into the real line; i.e., $U\colon \mathbb{R}^M\to\mathbb{R}$.

The planning bureau seeks the program $P^*=(x^*,y^*)$ that solves the planning problem:

$$\max_{x,y} U(x) \text{ subject to } x\in X$$

$$(1)\qquad y_i\in Y_i \quad i=1,\dots,N$$

$$x-\sum_{i=1}^{N} y_i\leqslant W$$

II. A Class of Planning Procedures

The class of planning procedures I consider are those that choose feasible solutions and coverage to an optimal plan (either finitely or in the limit). I therefore eliminate from consideration those procedures that ensure ε-convergence, that is, a plan "close" to the optimal plan.

To reiterate, let us assume that the planning board knows its own utility function $U(x)$ as well as the set of feasible final

[3] This model is the one developed by Malinvaud.

[4] By "know" I mean that the firm has the capability of calculating the coordinates of any point in its production set; for example, given a price vector, it can determine the vector in its production set with maximum value.

consumption vectors X and the vector W of initial availabilities of all goods. Each firm knows its own feasible production set Y_i, but not the sets of the other firms; the firms do not know the utility function of the planning board nor the set X. If the planning board knew all the production sets Y_i, it could solve its planning problem (1) directly.

The purpose of the planning procedure is thus to obtain information from the firms about their feasible production sets. We will assume that the planning board knows that the Y_i are all members of the same subset $\mathbf{H}$ of sets in the commodity space $\mathbb{R}^M$. For example, $\mathbf{H}$ may be the set of convex subsets of $\mathbb{R}^M$, so that the planning board knows that all the firms' production-possibilities sets are convex. If the firms choose to misrepresent their production sets (i.e., cheat), they must do so in a way that is not detectable by the planning board.

More formally, if Y_i is firm i's true production set, let $\mathbf{Y}_i \subset \mathbb{R}^M$ denote firm i's *reported* production set. Firm i will be said to be "telling the truth" if its reported production set is its true production, i.e., if $\mathbf{Y}_i = Y_i$.

At each stage or iteration of the planning process, the planning bureau receives messages from the firms in response to its messages. The set of possible firm messages and planning board messages are denoted by $\mathbf{M}$ and $\mathbf{P}$, respectively. For example, the firms might be reporting production vectors, so that $\mathbf{M} = \mathbb{R}^M$, while the planning board might be sending price vectors, so that $\mathbf{P}$ is the unit simplex in $\mathbb{R}^M$. Let $p_i^t \in \mathbf{P}$ be the message sent to firm i at iteration t; let $m_i^t \in \mathbf{M}$ be the reply of firm i.

DEFINITION 1: *The sequence $\{m_i^t\}_{t=1}^{\infty}$ will be said to be consistent with a set $\mathbf{Y}_i$ given $\{p_i^t\}_{t=1}^{\infty}$ if $\{m_i^t\}_{t=1}^{\infty}$ are the messages that would result from following the planning board's directives when $\mathbf{Y}_i \subset \mathbb{R}^M$ is the firm's production set.*[5]

[5]Since it may not be decidable in general whether or not $\{m_i^t\}_{t=1}^{\infty}$ is consistent with a particular set $\mathbf{Y}_i \subset \mathbb{R}^M$, I am implicitly excluding planning procedures for which this is not decidable.

Thus, a sequence of messages $\{m_i^t\}_{t=1}^{\infty}$ is consistent with a set in $\mathbf{H}$ if $\mathbf{Y}_i \in \mathbf{H}$ exists satisfying Definition 1.

The messages sent to the planning board by each firm depend upon the production set $\mathbf{Y}_i$ the firm chooses to report as well as the board's messages $\{p_i^t\}_{t=1}^{\infty}$. The planning board's messages to each division are assumed to be nonrandom, that is, functionally dependent on the reported production sets of *all* the firms. We can express each firm's messages as a function of *all* the reported production sets:

$$m_i^t = m_i^t(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$$

All the messages received by the planning board from every firm at every iteration will be denoted by $m^{\infty} \equiv (m_1^1, \ldots, m_N^1, m_1^2, \ldots, m_N^2, \ldots)$.

Define $(x(m^{\infty}), y(m^{\infty}))$ to be the vector of final consumption and firm production plans chosen by the planning procedure when the firms send messages m^{∞} to the planning board:

DEFINITION 2: *A planning procedure will be called admissible if when the reported production sets $\mathbf{Y}_i \in \mathbf{H}$ for $i = 1, \ldots, N$, $(x(m^{\infty}(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)), y(m^{\infty}(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)))$ solves the planning problem* (1) *where Y_i has been replaced by $\mathbf{Y}_i$ for all i.*

Given this definition of an admissible procedure, if the firms all report their true production sets Y_i, the procedure finds the solution to the planning board's problem (1).

III. Definition of Firm Evaluations

Given the definition of an admissible planning procedure, we can define the enforcement rules. The enforcement rules are defined implicitly by an N-tuple of evaluation measures, one for each firm. When the evaluation measures are appropriately defined, maximization by a firm of its individual measure "enforces" the following of the planning board's directives. It should be pointed out that these evaluations provide index numbers on which individual firm rewards can be based; they do not represent real resource flows. For example, the reward to one firm might be in the form of paid vacation days, while another firm might be rewarded with salary increases.

DEFINITION 3: *Firm i's evaluation measure is a function* $E_i: \mathbf{M}^\infty \to \mathbb{R}$ *defined as follows:*[6]

$$E_i[m^\infty] \equiv \begin{cases} U(x(m^\infty)) \text{ if } \{m_i^t\}_{t=1}^\infty \text{ is consistent with a set in } \mathbf{H} \text{ given } \{p_i^t\}_{t=1}^\infty \text{ and } y_i(m^\infty) \text{ is feasible} \\ \inf_{x \in X} U(x) - 1 \text{ otherwise} \end{cases}$$

These evaluations do not preclude firm cheating directly. The firm must appear to be reporting its true production set; in addition if $y_i(m^\infty) \notin Y_i$ and thus not implementable, the planning board knows with certainty that the firm lied about its production set.

Payoff to each firm can be expressed as a function of the reported production sets:

$$W_i[\mathbf{Y}_1, \ldots, \mathbf{Y}_N] \equiv E_i[m^\infty(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)]$$

The problem for each firm is thus the choice of a reported production set that maximizes the firm's payoff.

If firm i knows only that the other firms are reporting production sets in $\mathbf{H}$ (possibly the true production sets, but not necessarily), firm i's best response is to report its true production set Y_i.

Formal (mathematical) statements of all theorems and corollaries as well as their proofs (when necessary) are given in the Appendix. In addition, although not stated in each theorem and corollary explicitly, the results are only shown to hold for the planning procedures and evaluation measures defined by Definitions 2 and 3, respectively.

[6] $\mathbf{M}^\infty \equiv \underset{t=1}{\overset{\infty}{X}} \mathbf{M}$

THEOREM 1: *Firm i maximizes its payoff by reporting its true production set if the other* $(N-1)$ *firms all report production sets in the set* $\mathbf{H}$.

The following is immediate from Theorem 1:

COROLLARY 1: *Firm i maximizes its payoff by reporting its true production set when every other firm is reporting its true production set; i.e.,* $(Y_1, \ldots, Y_N)$ *is a Nash equilibrium.*

Is telling the truth the only optimal strategy for firm i? The answer is no. Nevertheless, one can show that any production set $\mathbf{Y}_i^*$ that maximizes firm i's payoff for any possible (reported) production sets in $\mathbf{H}$ of the other firms must have a nonempty intersection with Y_i, i.e., $\mathbf{Y}_i^* \cap Y_i \neq \varnothing$.

THEOREM 2: *If firm i maximizes its payoff by reporting* $\mathbf{Y}_i^*$ *no matter what production sets in* $\mathbf{H}$ *the other* $(N-1)$ *firms report, then* $\mathbf{Y}_i^* \cap Y_i \neq \varnothing$.

Firm i will cheat only if cheating cannot be detected *and* does not reduce the firm's payoff if undetected. Therefore firm i will exclude points of its true production set only if there exist no production sets of the other firms that would lead the planning bureau to choose the excluded point given the set X and its constraints on feasible final consumption. Similarly, firm i will include infeasible points in its reported production set only if it is sure these infeasible points will never be chosen.

Calculation of an optimal production set to report (other than Y_i) requires knowledge of the set of feasible final consumption vectors X. Thus Y_i is the only reported production set that maximizes payoff and is totally independent of the reported production sets of the other firms, and the set X.

Intermixtures of Nash equilibria are generally not Nash equilibria; we would thus like to ensure that the firms have an incentive to arrive at a Nash equilibrium that leads to a decision that solves problem (1). The N-tuple $(Y_1, \ldots, Y_N)$ can be shown to dominate for every firm in the set of Nash equilibria; in addition, any such dominant Nash equilibrium leads to a decision solving problem (1).

THEOREM 3: *The payoff to each firm is at least as great at the Nash equilibrium* $(Y_1,\ldots,Y_N)$ *as it is at any other Nash equilibrium.*

THEOREM 4: *If the payoff to each firm under the Nash equilibrium* $(\mathbf{Y}_1^*,\ldots,\mathbf{Y}_N^*)$ *is the same as under the Nash equilibrium* $(Y_1,\ldots,Y_N)$, *i.e., when every firm reports its true production set, then*

$$\left(x(m^\infty(\mathbf{Y}_1^*,\ldots,\mathbf{Y}_N^*)),y(m^\infty(\mathbf{Y}_1^*,\ldots,\mathbf{Y}_N^*))\right)$$

solves the planning board's problem (1).

IV. A Procedure Implying Mathematical Programming at the Center (*MPC*)

Malinvaud defines three planning procedures for solving problem (1). The first is a tâtonnement procedure; the second is a procedure based on the Leontief-Samuelson technology; and the third is a procedure implying the use of mathematical programming by the planning bureau. It is this third procedure with which this section is concerned. Malinvaud makes the following assumptions about $U(\cdot)$, X, the sets Y_i, and the planning board's a priori knowledge:

(a) The set X is closed and convex. There exists an $\bar{x}\in X$ such that $x\geqslant\bar{x}$ for all $x\in X$. The utility function $U(x)$ is concave and continuous.

(b) The sets Y_i, $i=1,\ldots,N$, are closed, bounded and convex.

(c) The planning board knows a feasible solution to the planning problem (1).

In this procedure prices are used to decompose problem (1). At each stage of the process, the planning bureau sends the firm a (normalized) price vector p^{t-1} for their outputs. Each firm then maximizes the value of its output at the price vector, that is, each firm solves a problem of the following form:

$$\text{(2)}\qquad \max_{y_i} p^{t-1}y_i \qquad \text{subject to } y_i\in Y_i$$

The firm then sends to the planning bureau the solution y_i^{t-1} to the firm's problem. The bureau uses this maximizer along with the maximizers received at previous iterations to approximate Y_i, the firm's feasible production set. The approximated production set Y_i^t for firm i at iteration t is the set of all convex combinations of maximizers received from firm i at all iterations, including the current one:

$$Y_i^t\equiv\left\{y_i\,|\,y_i=\sum_{\tau=1}^{t-1}\lambda_{i\tau}y_i^\tau,\ \sum_{\tau=0}^{t-1}\lambda_{i\tau}=1,\ \lambda_{i\tau}\geqslant 0,\ \tau=0,1,\ldots,t-1\right\}$$

The planning bureau then solves for what the optimal vectors of final consumption and net outputs would be if the firms' production sets were actually the sets Y_i^t, $i=1,\ldots,N$, i.e., the planning bureau solves problem (1) with Y_i replaced by Y_i^t for all i. Let $(x^t,y^{(t)})$ solve the planning board's problem at iteration t. Revised prices p^t are any (normalized) M-vector such that $p^t\geqslant 0$ and satisfying the following three conditions:

1) $p^t x > p^t x^t$ for all $x\in X$ such that $U(x)>U(x^t)$

2) $p^t y_i^{(t)}\leqslant p^t y_i$ for all $y_i\in Y_i^t$, for all $i=1,\ldots,N$

3) $p_k^t=0$ for all k such that

$$x_k^t-\sum_{i=1}^N y_{ki}^{(t)}<W_k$$

The planning bureau's decision is

$$(x(m^\infty),y(m^\infty))\equiv\left(\lim_{t\to\infty}x^t,\ \lim_{t\to\infty}y^{(t)}\right)$$

Since the planning procedure converges to a solution of problem (1) (see Malinvaud) it is straightforward to verify its admissibility. If at each stage each division i solves a problem like (2) where Y_i is replaced by $\mathbf{Y}_i$, and where $\mathbf{Y}_i$ is convex, then it is immediate by admissibility that

$$\left(x(m^\infty(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)),\ y(m^\infty(\mathbf{Y}_1,\ldots,\mathbf{Y}_N))\right)$$

solves problem (1) where Y_i has been replaced by $\mathbf{Y}_\mathbf{i}$ for all i.

In order to apply the evaluation measures defined in Section III, and therefore the results of that section, it must be shown that the planning bureau can judge whether or not a sequence of vectors $\{y_i^t\}_{t=1}^{\infty}$ is consistent with some convex set $\mathbf{Y}_\mathbf{i}$. Since the firm is maximizing a linear function over a convex set, the maximizer y_i must be an extreme point of $\mathbf{Y}_\mathbf{i}$. Therefore a sequence $\{y_i^t\}_{t=1}^{\infty}$ will be consistent with a convex set $\mathbf{Y}_\mathbf{i}$ if and only if the following is true for all t:

$$y_i^t \notin (\text{interior } Y_i^t)$$

If $\mathbf{Y}_\mathbf{i}$ were a convex polyhedral set, then y_i^t could certainly be a point on one of the faces of Y_i^t; i.e., $y_i^t \in$(boundary Y_i^t).

The evaluation measure for each firm is then:

$$E_i[m^\infty] = \begin{cases} U(x(m^\infty)) & \text{if } y_i^t \notin (\text{interior } Y_i^t) \\ & \text{for all } t; \\ & \text{and } \lim\limits_{t\to\infty} y_i^{(t)} \in Y_i \\ \inf\limits_{x\in X} U(x) - 1 & \text{otherwise} \end{cases}$$

If the firms in Malinvaud's economy are evaluated in the above manner, there is reason to expect that the rules of *MPC* will be followed, based on the results in Section III. These results state that the N-tuple of true production sets $(Y_1,\ldots,Y_N)$ constitute a Nash equilibrium that dominates in the set of Nash equilibria. Moreover, this is the only Nash equilibrium known a priori to all the firms in the economy and that is independent of the set X of feasible final consumption vectors.

Appendix

All theorems and corollaries hold for all admissible planning procedures and evaluation measures defined by Definitions 2 and 3, respectively.

THEOREM 1: *For all* $\mathbf{Y}_\mathbf{j} \in \mathbf{H}, j\neq i$, *if* $\mathbf{Y}_\mathbf{i}^* = Y_i$ *then* $W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_\mathbf{i}^*,\ldots,\mathbf{Y}_N] \geqslant W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_N]$ *for all* $\mathbf{Y}_\mathbf{i} \subset \mathbb{R}^M$.

PROOF:

By admissibility of the planning procedure, $(x(m^\infty(\mathbf{Y}_1,\ldots,Y_i,\ldots,\mathbf{Y}_N)),\ y(m^\infty(\mathbf{Y}_1,\ldots,Y_i,\ldots,\mathbf{Y}_N)))$ solves

$$\max_{x,y} U(x)$$

subject to $x\in X$; $y_i \in Y_i$; $y_j \in \mathbf{Y}_\mathbf{j}$ $j\neq i$; and $x-\sum_{i=1}^N y_i \leqslant W$.

Let $\mathbf{Y}_\mathbf{i}$ be any other subset of $\mathbb{R}^M$ (not necessarily convex), and $(x(m^\infty(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)), y(m^\infty(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)))$ the vector of final consumption and production chosen by the planning board when firm i reports its production set as $\mathbf{Y}_\mathbf{i}$. One of the following must hold:

(i) $y_i(m^\infty(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)) \in Y_i$. The production plan chosen when firm i reports $\mathbf{Y}_\mathbf{i}$ is still available if the firm reports Y_i, in addition to production plans that were not previously available. It is immediate that $U(x(m^\infty(\mathbf{Y}_1,\ldots,Y_i,\ldots,\mathbf{Y}_N))) \geqslant U(x(m^\infty(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)))$ and $W_i[\mathbf{Y}_1,\ldots,Y_i,\ldots,\mathbf{Y}_N] \geqslant W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_N]$.

(ii) $y_i(m^\infty(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)) \notin Y_i$. Then $W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_N] = \inf_{x\in X} U(x) - 1$ and $W_i[\mathbf{Y}_1,\ldots,Y_i,\ldots,\mathbf{Y}_N] \succ W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_N]$.

COROLLARY 1: *The N-tuple* $(Y_1,\ldots,Y_N)$ *is a Nash equilibrium, i.e., for all* $i=1,\ldots,N$, $W_i[Y_1,\ldots,Y_N] \geqslant W_i[Y_1,\ldots,\mathbf{Y}_\mathbf{i},\ldots,Y_N]$ *for all* $\mathbf{Y}_\mathbf{i} \subset \mathbb{R}^M$.

PROOF:

Immediate from Theorem 1.

THEOREM 2: *If for all* $\mathbf{Y}_\mathbf{j} \in \mathbf{H}$, $j\neq i$, $W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_\mathbf{i}^*,\ldots,\mathbf{Y}_N] \geqslant W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_N]$ *for all* $\mathbf{Y}_\mathbf{i} \subset \mathbb{R}^M$, *then* $\mathbf{Y}_\mathbf{i}^* \cap Y_i \neq \phi$.

PROOF:

Assume $\mathbf{Y}_\mathbf{i}^* \cap Y_i = \phi$. Then $y_i(m^\infty(\mathbf{Y}_1,\ldots,\mathbf{Y}_\mathbf{i}^*,\ldots,\mathbf{Y}_N)) \notin Y_i$ for all convex $\mathbf{Y}_\mathbf{j}$, $j\neq i$, and $W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_\mathbf{i}^*,\ldots,\mathbf{Y}_N] = \inf_{x\in X} U(x) - 1$ for some convex $\mathbf{Y}_\mathbf{j}$, $j\neq i$. Therefore $W_i[\mathbf{Y}_1,\ldots,Y_i,\ldots,\mathbf{Y}_N] \succ W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_\mathbf{i}^*,\ldots,\mathbf{Y}_N]$ which is a contradiction (from Theorem 1).

THEOREM 3: *For every Nash equilibrium* $(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)$, $W_i[Y_1,\ldots,Y_N] \geqslant W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_N]$ *for all* $i=1,\ldots,N$.

PROOF:

If $P=(x(m^{\infty}(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)),\ y(m^{\infty}(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)))$ is chosen when the firms report $(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)$, when the firms report $\mathbf{Y}'_1=\{y_i(m^{\infty}(\mathbf{Y}_1,\ldots,\mathbf{Y}_N))\}$, the planning board will again choose P. Therefore, $U(x(m^{\infty}(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)))=U(x(m^{\infty}(\mathbf{Y}'_1,\ldots,\mathbf{Y}'_N)))$. Since $(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)$ is a Nash equilibrium, $y_i(m^{\infty}(\mathbf{Y}_1,\ldots,\mathbf{Y}_N))\in Y_i$ for all i. Thus, $\mathbf{Y}'_i\subset Y_i$ for all i and $U(x(m^{\infty}(\mathbf{Y}'_1,\ldots,\mathbf{Y}'_N)))\leqslant U(x(m^{\infty}(Y_i,\ldots,Y_N)))$.

We therefore have $W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_N]\equiv U(x(m^{\infty}(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)))=U(x(m^{\infty}(\mathbf{Y}'_1,\ldots,\mathbf{Y}'_N)))\leqslant U(x(m^{\infty}(Y_1,\ldots,Y_N)))\equiv W_i[Y_1,\ldots,Y_N]$

THEOREM 4: *If* $(\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N)$ *is a Nash equilibrium and for all* i, $W_i[Y_1,\ldots,Y_N]=W_i[\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N]$, *then* $P^*=(x(m^{\infty}(\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N)),\ y(m^{\infty}(\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N)))$ *solves problem* (1).

PROOF:

(i) Since $(\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N)$ is a Nash equilibrium, $P^*=(x(m^{\infty}(\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N)),\ Y(m^{\infty}(\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N)))\in X\times Y_1\times\ldots\times Y_N$

(ii) By admissibility, $(x(m^{\infty}(Y_1,\ldots,Y_N)),\ y(m^{\infty}(Y_1,\ldots,Y_N)))$ solves (1). By assumption, $U(x(m^{\infty}(Y_1,\ldots,Y_N)))\equiv W_i[Y_1,\ldots,Y_N]=W_i[\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N]\equiv U(x(m^{\infty}(\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_N)))$. From (i), P^* is feasible. Thus, P^* solves (1).

REFERENCES

K. Arrow, "Control in Large Organizations," *Manage. Sci.*, Apr. 1964, *10*, 397–408.

______ **and L. Hurwicz**, "Decentralization and Computation in Resource Allocation," in Ralph Pfouts, ed., *Essays in Economics and Econometrics*, Chapel Hill 1960, 34–104.

S. Cohen, "Incentives, Iterative Communication and Organizational Control," *J. Econ. Theory*, Feb. 1980, *22*, 37–55..

______, "Incentive Compatible Control of the Multidivisional Firm with Iterative Communication," unpublished doctoral dissertation, Northwestern Univ. 1977.

J. Drèze and D. de la Vallee Poussin, "A Tâtonnement Process for Public Goods," *Rev. Econ. Stud.*, Apr. 1971, *38*, 133–50.

T. Groves, "Incentive Compatible Control of Decentralized Organizations," work. paper no. 166, Center Math. Stud. Econ. Manage. Sci., Northwestern Univ., Aug. 1975.

Geoffrey Heal, *The Theory of Economic Planning*, Amsterdam 1973.

L. Hurwicz, "The Design of Mechanisms for Resource Allocation," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 1–30.

P. Jennergren, "Studies in the Mathematical Theory of Decentralized Resource Allocation," unpublished doctoral dissertation, Stanford Univ. 1971.

M. Loeb and W. Magat, "Success Indicators in the Soviet Union: The Problem of Incentives and Efficient Allocation," *Amer. Econ. Rev.*, Mar. 1978, *68*, 173–81.

E. Malinvaud, "Decentralized Procedures for Planning," in his and M. O. L. Bacharach, eds, *Activity Analysis in the Theory of Growth and Planning*, London 1967, 170–208.

J. Roberts, "Incentives in Planning Procedures for the Provision of Public Goods," *Rev. Econ. Stud.*, forthcoming.

F. Schoumaker, "Best Reply Strategies in the Champsaur-Drèze-Henry Process," unpublished manuscript, Northwestern Univ., Nov. 1977.

NATIONAL SCIENCE FOUNDATION
Washington, D.C. 20550

FINAL PROJECT REPORT
NSF FORM 98A

PLEASE READ INSTRUCTIONS ON REVERSE BEFORE COMPLETING

PART I—PROJECT IDENTIFICATION INFORMATION

1. Institution and Address	2. NSF Program	3. NSF Award Number
Georgia Institute of Technology 225 North Avenue Atlanta, Georgia 30332	Economics	SOC 78-05770
	4. Award Period From 7/1/78 To 12/31/80	5. Cumulative Award Amount $44,448

6. Project Title

Incentives, Organizational Design and Iterative Communication

PART II—SUMMARY OF COMPLETED PROJECT *(FOR PUBLIC USE)*

One of the major considerations in organizational design is the enforcement procedures which induce members to follow announced operating rules. Prior to this project, research on the specification of enforcement rules had suppressed the effects of iterative communication between the "central decision maker" and members of the organization. The objective of this project was to build sequemtial decision process models of both intermittant communication and enforcement. The expectation was that the results would contribute to economics and, perhaps, to the literature in sequential games, accounting, and organizational behavior. A principal tool of analysis was expected to be the extant theory of sequential games.

The actual results of the project fell short of the aspirations. Numerous sequential mathematical models of both intermittant communication and enforcement were constructed, but none of them yielded to analysis. Therefore, the positive results of the project are of two kinds. First, further results were obtained in the case where only communication is iterative. Second, numerous results were obtained for the optimization of stochastic models, properties of general sequential game models, and the application of these general results to economic models.

PART III—TECHNICAL INFORMATION *(FOR PROGRAM MANAGEMENT USES)*

1. ITEM *(Check appropriate blocks)*	NONE	ATTACHED	PREVIOUSLY FURNISHED	TO BE FURNISHED SEPARATELY TO PROGRAM: Check (√)	Approx. Date
a. Abstracts of Theses					
b. Publication Citations		XX		see attached	
c. Data on Scientific Collaborators					
d. Information on Inventions					
e. Technical Description of Project and Results					
f. Other *(specify)*					

2. Principal Investigator/Project Director Name *(Typed)*	3. Principal Investigator/Project Director Signature	4. Date
Susan I. Cohen Matthew J. Sobel	Susan I. Cohen Matthew J. Sobel	3/31/81

NSF Form 98A (5-78) Supersedes All Previous Editions

Form Approved OMB No. 99R0013

Publication Citations
SOC 78-05770

S.I. Cohen, Incentives,Iterative Communication and Organizational Control, Journal of Economic Theory, Vol. 22, No. 1, February 1980.*

S.I. Cohen, Incentives and the Choice of Optimal Plans, American Economic Review, Vol. 70, No. 4, September 1980.*

S.I. Cohen, Incentive Properties of Iterative Decomposition Algorithms, accepted for publication, Management Science.*

S.I. Cohen, Truth-Telling, Dominant Strategies and Iterative Groves Mechanisms, in review, Journal of Economic Theory.*

R. Mendelssohn and M.J. Sobel, Capital Accumulation and the Optimization of Renewable Resource Models, Journal of Economic Theory, Vol. 23, pp. 243-260, 1980.*

M.J. Sobel and M. Shubik, Stochastic Games, Oligopoly Theory, and Competitive Resource Allocation, pp. 89-104 in Dynamic Optimization and Mathematical Economics, P.-T. Liu, Ed., Plenum Publishing Corp., New York, 1980.*

M.J. Sobel, Ordinal Sequential Games, Economies et Societes, to appear, 1981.*

M.J. Sobel, Myopic Solutions to MDPs and Stochastic Games, Operations Research, to appear, 1981.*

D.P. Heyman and M.J. Sobel, Stochastic Models in Operations Research, Vol. 2 (Stochastic Optimization), Chapter VIII (Sequential Games) and Section 3 of Chapter VII (Resource Management), McGraw-Hill, New York, 1982.**

M. Shubik and M.J. Sobel, Market Structure and Behavior, Vol. 2 (Vol. 1 has been published by Harvard University Press, coauthored with R. Levitan and M. Shubik, and that press has first rights of refusal on Vol. 2 which is expected to be completed during 1982).***

M.J. Sobel, Stochastic Fishery Games with Myopic Equilibria, in The Economics of Renewable Resource Management, L.J. Mirman and D.J. Spulber, Eds., Elsevier-North Holland Publishing Co., New York, 1981 (to appear).*

M.J. Sobel, Homogeneous Markov Decision Processes, Working Paper MS-80-10, College of Management, Georgia Institute of Technology, Atlanta, 1980.*

* attached
** to be sent April 15, 1981
*** to be sent September 30, 1982

ORDINAL SEQUENTIAL GAMES*

Matthew J. Sobel

Georgia Institute of Technology

ABSTRACT

The transitive completion of an ordinal sequential game is shown to possess a solution only if it has one amongst stationary policies. The kinds of solutions investigated are equilibrium points and the core (including Pareto optima). Players' preferences are assumed to be persistent and countably transitive with the second assumption being akin to discounting. The preferences need be neither transitive, reflexive, nor complete. The results pertain also to one-player sequential decision processes with multi-attribute utility functions.

KEY WORDS

Stochastic games, sequential decision, equilibrium point, Pareto optimum, ordinal preferences, core.

---

*Partially supported by NSF grant ENG76-15559 at Yale University and NSF grant SOC78-05770 at Georgia Institute of Technology.

ORDINAL SEQUENTIAL GAMES

Matthew J. Sobel

Georgia Institute of Technology

1. Introduction

Since Shapley's seminal article [7], research on sequential games has demonstrated existence of an equilibrium point in stationary policies under ever more general conditions. This research, recently surveyed in [6], concerns discounted and average return criteria. The present article addresses the discounted case. The central result is that the transitive completion of a game has an equilibrium point (respectively, a Pareto optimum) in stationary policies if the original game has *any* equilibrium point (Pareto optimum).

The arguments here do not entail assumptions of cardinal utility. Instead, each player is assumed to possess a countably transitive preference structure that exhibits temporal persistence. Countable transitivity is akin to discounting and temporal persistence has been called "consistent choice." The consequences of the latter assumption are examined in [3] and the other assumptions are discussed briefly in [8]. The preference structures need not be transitive, reflexive, or complete.

It is a pleasure to correct an oversight in [8] and acknowledge the contribution of Denardo's thesis [1]. Although the preferences in [1] are cardinal, the assumptions are closely related to mine here and in [8]. In particular, countable transitivity plays the same role as Denardo's "convergence condition."

In outline, §2 specifies the deterministic structure to which the results derived in §3 first apply. Games with the following generalizations are discussed in §4: stochastic transitions, pure strategy solutions, learning, and nonstationary transition probabilities.

The specific results in §3 apply to the core, hence Pareto optima, as well as equilibrium points. Therefore, they pertain to one-person sequential decision processes with a multi-attribute utility function.

2. A Sequential Decision Process

Let S be a nonempty set of *states* and A(s) a nonempty set of *actions* for each $s \varepsilon S$. Let $C = \{(s,a): a \varepsilon A(s), s \varepsilon S\}$, $\Delta = \times_{s \varepsilon S} A(s)$, and $Y = \times_{t=1}^{\infty} \Delta$. Elements of Δ are single *stage decision rules* and elements of Y are *Markov policies*. Transitions from state to state are determined by a mapping M from C to S. A *posterity* is a feasible sequence of successive states and actions. The set Φ_s of all posterities with initial state s is

$$\Phi_s = \{(s_1,a_1,s_2,a_2, \ldots): s_1 = s, (s_n,a_n) \varepsilon C \text{ and } s_{n+1} = M(s_n,a_n) \text{ for all } n\}$$

A Markov policy $\pi = (\delta_1,\delta_2, \ldots)$ is a *stationary policy* if $\delta_1 = \delta_2 = \ldots = \delta$ for some $\delta \varepsilon \Delta$. Then π is written as δ^{∞}.

Let G be a nonempty set and $B \subseteq G \times G$. If $(b,b) \varepsilon B$ for all $b \varepsilon G$ then B is *reflexive*. If $(b,c) \varepsilon B$ and $(c,d) \varepsilon B \Rightarrow (b,d) \varepsilon B$ then B is *transitive*. For all $(b,c) \varepsilon B \times B$, if $(b,c) \varepsilon G$, or $(c,b) \varepsilon G$, or both $(b,c) \varepsilon G$ and $(c,b) \varepsilon G$, then B is *complete*. If B is complete, then it is reflexive. Let f be a mapping from G to G. Then f is *isotone* (with B) if $(b,c) \varepsilon B \Rightarrow [f(b),f(c)] \varepsilon B$. An element $b \varepsilon G$ is B-*maximal* if $(b,c) \varepsilon B$ for all $c \varepsilon G$.

For each $\gamma \varepsilon \Delta$, let T_γ be the mapping from Y to Y given by $T_\gamma \pi = (\gamma,\delta_1,\delta_2, \ldots)$ where $\pi = (\delta_1,\delta_2, \ldots)$. Thus $T_\gamma \pi$ defers the use of π for one period during which the single period decision rule γ is used.

The results in [8] depend on three axioms which concern a binary relation

$B \subseteq Y \times Y$. We write $\pi \geq \xi \iff (\pi,\xi)\varepsilon B \iff \xi \leq \pi$. The axioms are:

$$(\geq, Y) \textit{ is reflexive and transitive}; \tag{1}$$

$$\textit{for each } \gamma\varepsilon\Delta,\ T_\gamma \textit{ is isotone on } (\geq, Y); \tag{2}$$

$$\left.\begin{array}{l} \textit{for each } \pi = (\delta_1, \delta_2, \ldots)\varepsilon Y \textit{ and } \xi\varepsilon Y, \\ \qquad \xi \geq T_{\delta_1} T_{\delta_2} \ldots T_{\delta_k}\xi \textit{ for all } K < \infty \implies \xi \geq \pi \\ \quad T_{\delta_1} T_{\delta_2} \ldots T_{\delta_k}\xi \geq \xi \textit{ for all } K < \infty \implies \pi \geq \xi. \end{array}\right\} \tag{3}$$

The axioms are called *rationality, consistent choice* or *temporal persistence* or *stationarity,* and *countable transitivity, respectively.*

THEOREM 1 *Suppose axioms* (1), (2), *and* (3) *are valid.*

A. $T_\delta\pi \geq \pi \implies \delta^\infty \geq \pi$ (Theorem 1 in [8]).

B. *If there exists a* $\geq$*-maximal* $\pi\varepsilon Y$ *then there exists* $\delta\varepsilon\Delta$ *such that* δ^∞ *is* $\geq$*-maximal* (Theorem 2 in [8]).

C. $\pi\varepsilon Y$ *is* $\geq$*-maximal* $\iff \pi \geq T_\delta\pi$ *for all* $\delta\varepsilon\Delta$ (Theorems 3 and 4 in [8]).

D. *If* $|C| < \infty$ *then there exists a* $\geq$*-maximal* δ^∞ (Corollary in [8]).

E. *If* Δ *is compact in a topology for which* $\{\gamma\colon \gamma^\infty \geq \delta^\infty, \gamma\varepsilon\Delta\}$ *is closed for every* $\delta\varepsilon\Delta$, *then there exists a* $\geq$*-maximal* δ^∞ (Theorem 6 in [8]).

Part B can be used to restate the => portion of part C:

F. *For every* $\gamma\varepsilon\Delta$, *either* γ^∞ *is* $\geq$*-maximal or there is another* $\delta\varepsilon\Delta$ *such that* $\delta^\infty \geq \gamma^\infty$.

The proof of F in [8] is constructive and is, essentially, the policy improvement algorithm. Therefore, under the assumptions of Theorem 3 below, that algorithm can be applied to the transitive completions (specified in §3) of the orderings $(\geq_e, Y)$ and $(\geq_p, Y)$ defined below. We do not explore the algorithm's performance in this paper. The theorem is valid too for stochastic sequential decision processes (§4).

The binary relation $(\geq, Y)$ can arise in several ways. In a one-player model, the relation is induced by a family of binary relations $\theta_s \subseteq \Phi_s \times \Phi_s$ for each $s \in S$ with the interpretation $(p,p') \in \theta_s \iff$ posterity p is at least as desireable as posterity p'. For $s \in S$ and $\pi = (\delta_1, \delta_2, \ldots) \in Y$, let $p_s(\pi)$ denote $(s, a_1, s_2, a_2, \ldots) \in \Phi_s$ where $a_n = \delta_n(s_n)$ for all n. In [8],

$$\pi \geq \gamma \iff [p_s(\pi),\ p_s(\gamma)] \in \Phi_s \text{ for all } s \in S.$$

An equivalence between $(\geq, Y)$ and $\{\theta_s : s \in S\}$ is detailed in [4] and [5, §XVII.3]. In particular, if $\{\theta_s : s \in S\}$ possess rationality, temporal persistence, and countable transitivity then $(\geq, Y)$ inherits these properties.

Multiple player models can induce $(\geq, Y)$. Let I be a nonempty set of *players* and A_s^i a nonempty set of actions available to player i when the state is s. Let $A(s) = \times_{i \in I} A_s^i$. Then define Δ, Y, C. and M as previously done. A *coalition* is a nonempty subset of players. Let Ω, the set of latent coalitions, be a nonempty collection of nonempty subsets of I. For each $\omega \in \Omega$ and $s \in S$, let $\theta_s^\omega \subseteq \Phi_s \times \Phi_s$ indicate the coalition's preferences concerning posterities with $s_1 = s$. This is a *sequential game* model.

It is convenient to label the i-*th* player's portion of $\pi \in Y = \times_{t=1}^{\infty} \times_{s \in S} \times_{i \in I} A_s^i$ as π^i and the portion of π due to all the remaining players as π^{-i}. With this notation, binary relations underlying definitions of equilibrium point and Pareto optimum, respectively, are

$$\left.\begin{array}{l}\pi \geq_e \xi \iff [p_s(\pi),\ p_s(\xi^i, \pi^{-i})] \in \theta_s^i \\ \qquad\qquad \textit{for all } s \in S \textit{ and } i \in I;\end{array}\right\} \quad (4)$$

$$\left.\begin{array}{l}\pi \geq_p \xi \iff \text{either } [p_s(\pi), p_s(\xi)] \in \theta_s^\omega \text{ for all } s \in S,\ \omega \in \Omega \\ \qquad \text{or there exist } (s,\omega) \text{ and } (u,j) \text{ such that} \\ \qquad [p_s(\pi), p_s(\xi)] \notin \theta_s^\omega \text{ and } [p_u(\xi), p_s(\pi)] \notin \theta_u^j.\end{array}\right\} \quad (5)$$

DEFINITION. π *is an* equilibrium point $\iff$ π *is* $\geq_e$*-maximal.* π *is in the* core $\iff$ π *is* $\geq_p$*-maximal.*

When $\Omega = I$, elements in the core are sometimes called Pareto optima.

The binary relation $(\geq, Y)$ can arise in a manner that mixes the one-player and multiple player models. Consider a single decision maker with Δ, Y, C, and M defined as in the one-player model. Let I denote a nonempty set of *criteria*. If $|I| > 1$ this is a *multi-criterion* model. For each $s \in S$ and $i \in I$, let $\theta_s^i \subseteq \Phi_s \times \Phi_s$ indicate preferences with respect to the i-*th* criterion if $s_1 = s$. Let (5) define $\geq_p$ with $\Omega = I$. Then a $\geq_p$-maximal policy is called *undominated, efficient,* and a *Pareto optimum* in various literatures.

If $\{\theta_s^\omega : s \in S,\ \omega \in \Omega\}$ has rationality, temporal persistence, and countable transitivity, $(\geq_e, Y)$ and $(\geq_p, Y)$ are *not* necessarily transitive. Hence, Theorem 1 cannot straightforwardly be applied to sequential games.

The difficulty has nothing to do with the dynamical nature of the process. Consider the following bimatrix game.

0, 0 r	0, -1 s	0, 1 t
-1, 0 u	0, 0 v	0, -1 w
1, 0 x	-1, 0 y	0,0 z

Then r $\geq_e$ v and v $\geq_e$ z but r $\not\geq_e$ z. Consider the following bimatrix game.

0, 1 w	3, 0 x
2, 2 y	0, 0 z

Then w $\geq_p$ x and x $\geq_p$ y but w $\not\geq_p$ y.

3. Solutions in Stationary Policies

The following resolution of the intransitivity of $(\geq_e, Y)$ and $(\geq_p, Y)$ was provoked by Whitt's paper [9] stimulated by Fishburn [2]. Let G be a nonempty set and $B \subseteq G \times G$. Using Whitt's notation, we write $xRy \Leftrightarrow (x,y)\varepsilon B$, and $xPy \Leftrightarrow xRy$ and not yRx. An *inconsistency cycle* connecting x and y is a finite sequence $x_1, \ldots, x_n$ with $x_1 = x_n = x$ and $x_k = y$ for some k, $1 < k < n$, such that $x_i R x_{i+1}$ for all i and $x_i P x_{i+1}$ for some i. In the 2×2 bimatrix game above, w, x, y, w is an inconsistency cycle under the relation $\geq_p$. Define a relation R' on G via $xR'y \Leftrightarrow$ either xRy, or neither xRy nor yRx so x and y are unrelated under R. Define a relation R_c on G via $xR_cy \Leftrightarrow$ either xR'y or there is an inconsistency cycle connecting x and y under R'. The relation R_c is the *transitive completion* of R.

The following properties of R_c can be proven straightforwardly.

LEMMA 1. A. *R_c is complete and transitive.*

B. $xRy \Rightarrow xR_cy$.

C. $x\varepsilon G$ *is R-maximal* $\Rightarrow$ x *is* R_c*-maximal.*

Consider a relation $(\geq, Y)$ which need be neither complete, reflexive, nor transitive. Let $(\geq_c, Y)$ denote the relation constructed from $(\geq, Y)$ as (R_c, G) was constructed from (R,G).

THEOREM 2. *Suppose* $(\geq, Y)$ *is a relation for which* (2) *and* (3) *are valid.*

A. $T_\delta \pi \geq_c \pi \Rightarrow \delta^\infty \geq_c \pi$.

B. *If there exists a* $\geq$*-maximal* $\pi \varepsilon Y$ *then* π *is* $\geq_c$*- maximal and there exists a* $\geq_c$*-maximal* δ^∞.

C. *If* $\pi \varepsilon Y$ *is* $\geq$*-maximal then* $\pi \geq_c T_\delta \pi$ *for all* $\delta \varepsilon \Delta$.

D. *If* $|C| < \infty$ *then there exists a* $\geq_c$*-maximal* δ^∞.

E. *If* Δ *is compact in a topology for which* $\{\gamma: \gamma^\infty \geq_c \delta^\infty, \gamma \varepsilon \Delta\}$ *is closed for every* $\delta \varepsilon \Delta$, *then there exists a* $\geq_c$*-maximal* δ^∞.

F. *For every* $\gamma \varepsilon \Delta$, *either* γ^∞ *is* $\geq_c$*-maximal or there is another* $\delta \varepsilon \Delta$ *such that* $\delta^\infty \geq_c \gamma^\infty$.

PROOF. Parts A through F depend on part A of the lemma and on $\geq_c$ inheriting (2) and (3) from $\geq$. Part B of the lemma applies that $\geq_c$ inherits (2) and (3) from $\geq$. Each part of this theorem uses its corresponding latter part of Theorem 1. Also, parts B through F depend on part C of the lemma. $\square$

In order to apply Theorem 2 to the sequential game and multi-criterion models, it will be assumed that $\{\theta_s^\omega: \omega \varepsilon \Omega, s \varepsilon S\}$ is temporally persistent (2) and countably transitive (3). The task then will be to verify that $\geq_e$ and $\geq_p$ inherit these properties.

A state $s \varepsilon S$ is *reachable* if it can be reached from some state, i.e. if there exists $(u,a) \varepsilon C$ such that $s = M(u,a)$. The set S is called reachable if s is reachable for all $s \varepsilon S$.

We say that $\{\theta_s^\omega: s \varepsilon S, \omega \varepsilon \Omega\}$ is temporally persistent if $(u,a) \varepsilon C$, $M(u,a) = s$, and $(p,p') \varepsilon \theta_s^\omega \Rightarrow [(u,a,p),(u,a,p')] \varepsilon \theta_u^\omega$.

For $j = 0,1, \ldots$ let $p^j = (s, a_1^j, s_2^j, a_2^j, \ldots) \varepsilon \Phi_s$ with p^j matching p^0 up to s_j^j and a_j^j, namely $s_t^j = s_t^0$ and $a_t^j = a_t^0$, $t = 1, 2, \ldots, j$. We say that

$\{\theta_s^\omega\}$ is countably transitive if both

$$(p^{j+1},p^j)\varepsilon\theta_s^\omega \text{ for all } j \geq 1 \Rightarrow (p^0,p^j)\varepsilon\theta_s^\omega \text{ for all } j \geq 1;$$

$$(p^j,p^{j+1})\varepsilon\theta_s^\omega \text{ for all } j \geq 1 \Rightarrow (p^j,p^0)\varepsilon\theta_s^\omega \text{ for all } j \geq 1.$$

LEMMA 2. *Suppose $\{\theta_s^\omega\}$ is temporally persistent and countably transitive.*
A. $(\geq_e,Y)$ *satisfies* (2) *and* (3).
B. $(\geq_p,Y)$ *satisfies* (3). *If* S *is reachable, then* $(\geq_p,Y)$ *satisfies* (3).
PROOF. Countable transitivity of $\{\theta_s^\omega\}$ straightforwardly implies (3) for $\geq_e$ and $\geq_p$ defined by (4) and (5). Temporal persistence of $\{\theta_s^\omega\}$ directly implies (2) for $\geq_e$.

Suppose S is reachable. Let $(\succ,Y)$ be defined by $\pi \succ \xi \Leftrightarrow [p_s(\pi),p_s(\xi)]\varepsilon\theta_s^\omega$ for all $s\varepsilon S$ and $\omega\varepsilon\Omega$. Then temporal persistence of $\{\theta_s^\omega\}$ implies $(\succ,Y)$ satisfies (2). Isotonicity of $(\geq_p,Y)$ involves two cases. If $\pi \geq_p \xi$ and $\pi \succ \xi$ then (2) for $(\succ,Y)$ implies $T_\delta\pi \geq_p T_\delta\xi$. If $\pi \geq_p \xi$ but $\pi \succ \xi$ is false, then $\xi \geq_p \pi$ is true. Hence, by (5), there exist (s,i) and (u,j) such that $[p_s(\xi),\ p_s(\pi)]\not\varepsilon\theta_s^i$ and $[p_u(\pi),\ p_u(\xi)]\not\varepsilon\theta_u^j$. Therefore, temporal persistence of $\{\theta_s^\omega\}$ and the assumption that s and u are reachable imply existence of (v,a) and (x,b) such that $a\varepsilon A(v)$, $b\varepsilon A(x)$, $s = M(v,a)$, $u = M(x,b)$, and

$$([v,a,p_s(\xi)],\ [v,a,p_s(\pi)])\not\varepsilon\theta_v^i, \quad ([x,b,p_u(\pi)],\ [x,b,p_u(\xi)])\not\varepsilon\theta_x^j.$$

Therefore, $T_\delta\pi \geq_p T_\delta\xi$ (and $T_\delta\xi \geq_p T_\delta\pi$). □

THEOREM 3. *Suppose $\{\theta_s^\omega\}$ is temporally persistent and countably transitive.*

A. $(\geq_e, Y)$ *has properties* A *through* E *in Theorem* 2, *where* $\geq_c$ *refers to the transitive completion of* $\geq_e$.

B. *If* S *is reachable then* $(\geq_p, Y)$ *has properties* A *through* E *in Theorem* 2, *where* $\geq_c$ *refers to the transitive completion of* $\geq_p$.

Theorem 3 provides sufficient conditions for existence of an equilibrium point (respectively, subset of the core) among stationary policies if the set of equilibrium points (if the core) is nonempty.

4. Generalizations

Theorem 3 encompasses mixed strategies and processes in which states at successive stages comprise a discrete-time conditionally Markov process. The details in §5 of [8] will be outlined here.

Let X denote the set of states of the underlying stochastic process and assume that X is denumerable. Let Z_x^i be the denumerable set of actions available to player i in state x, S be the set of all probability measures on X, and Δ^i be the set of all (randomized) rules that specify an action in Z_x^i for each $x \in X$; define $\Delta = \times_{i \in I} \Delta^i$ and $Y = \times_{t=1}^{\infty} \Delta$. Then, $s_1, s_2, \ldots$ is the sequence of marginal distributions on X. The theorems are valid in this stochastic setting but a non-Markov policy would not induce conditionally independent successive states so the joint distribution of successive states could not be reconstructed from $s_1, s_2, \ldots$. Therefore, sufficiently arbitrary preference orderings could lead to Y being genuinely restrictive.

The state-to-state mapping M may be nonstationary without invalidating any result. That is, for each t, let S_t, $A_t(s)$, and M_t depend on t, and suppose $M_t(s,a) \in S_{t+1}$ for all $s \in S_t$ and $a \in A_t(s)$. Let $s_{t+1} = M_t(s_t, a_t)$. Say that $s \in S_{t+1}$ is reachable if there exists $v \in S_t$ and $a \in A_t(v)$ such that $s = M_t(v,a)$;

and S_{t+1} is reachable if s is reachable for all $s \in S_{t+1}$. For Lemma 2B and Theorem 3B, assume that S_t is reachable for all $t > 1$. Then a careful review of the proofs in [8] on which Theorem 1 (here) is based, and of the arguments in this paper, shows that the lemmas and theorems here remain valid. This extension admits some learning processes and nonstationary transition probabilities in the stochastic case. Therefore, existence of a stationary solution rests on stationarity of preference, i.e. postulate (2), rather than on stationarity in the underlying dynamics.

Acknowledgement

I am grateful to a referee who, in 1976, offered the first bimatrix game at the end of §2 as a counterexample to rash assertions in a different manuscript with the same title as this paper.

REFERENCES

[1] Denardo, E. V., Sequential Decision Processes, Ph. D. thesis, Northwestern University (1965).

[2] Fishburn, P. C. "Stochastic Dominance without Transitive Preferences," Management Science, 24 (1978), 1268-1277.

[3] Hammond, P. J., "Changing Tastes and Coherent Dynamic Choice," Review of Economic Studies, 43 (1976), 159-173.

[4] Henig, M., "Dynamic Programming with Returns in a Partial Ordered System," unpublished manuscript, Univ. of British Columbia (1979).

[5] Heyman, D. P. and M. J. Sobel, Stochastic Models in Operations Research, Volume 2, McGraw-Hill, New York, to appear.

[6] Parthasarthy, T. and M. Stern, "Markov Games--A Survey," in Differential Games and Control Theory II, edited by E. O. Roxin, P.-T. Liu, and R. L. Sternberg, Lecture Notes in Pure and Applied Mathematics, Marcel Dekker Co., New York (1977).

[7] Shapley, L., "Stochastic Games," Proc. National Academy of Sciences, U.S.A., 39 (1953), 1095-1100.

[8] Sobel, M. J., "Ordinal Dynamic Programming," Management Science, 21 (1975), 967-975.

[9] Whitt, W., "Comparing Probability Measures on a Set with an Intransitive Preference Relation," Management Science, 25 (1979), 505-511.

HOMOGENEOUS MARKOV DECISION PROCESSES

Matthew J. Sobel
Georgia Institute of Technology

Abstract

Suppose that each random variable ξ in a Markov decision process is replaced by $\lambda\xi$, $\lambda > 0$. Let s be a state and let $f(s,\lambda)$ denote the value of an optimal policy in the new problem if s is the initial state. Sufficient conditions are presented for $f(s,\lambda) = \lambda^m f(s/\lambda,1)$. The conditions are satisfied by several standard models.

Key Words

Markov decision process, dynamic programming, homogeneous, sensitivity analysis.

1. Introduction. Let S denote the set of *states*, A_s the set of *actions* which are feasible in state s, and $C \triangleq \{(s,a): a \in A_s,\ s \in S\}$. For each $(s,a) \in C$, let $\xi_{s,a}$ be a random variable with sample space Ω (the same for all $(s,a) \in C$). Let g be a real-valued function on $C \times \Omega$ and let h be a function from $C \times \Omega$ to S.

Consider dynamic programming recursions and equations such as

$$f_n(s) = \sup\{E[g(s,a,\xi_{s,a}) + f_{n-1}\{h(s,a)\xi_{s,a})\}]: a \in A_s\}, \qquad s \in S, \tag{1.1}$$

and

$$v(s) = \sup\{E[g(s,a,\xi_{s,a})] + \beta v\{h(s,a,\xi_{s,a})\}]: a \in A_s\}, \qquad s \in S. \tag{1.2}$$

The result below concerns the effect on $f_n(\cdot)$ and $v(\cdot)$ if $\xi_{s,a}$ is replaced by $\lambda\xi_{s,a}$, $\lambda > 0$. Accordingly, let

$$f(s,\lambda) \triangleq \sup\{E[g(s,a,\xi_{s,a}) + J\{h(s,a,\xi_{s,a}),\lambda\}]: a \in A_s\},\ s \in S,\ \lambda > 0, \tag{1.3}$$

where J is a real-valued function on $S \times (0,\infty)$. Suppose that the expectation in (1.3) always exists and that $f(\cdot,\cdot)$ is finite on $S \times (0,\infty)$.

The following section presents and discusses conditions sufficient for $f(s,\lambda) = \lambda^m f(s/\lambda), 1)$. This result is applied in §3 to an inventory model and to a fishery model.

2. Homogeneity. A set $X \subseteq \mathbb{R}^n$ is a *cone* if $\lambda > 0$ and $x \in X$ implies $\lambda x \in X$. If X is a cone, a function $w: X \to \mathbb{R}$ is *homogeneous of degree m* if $\lambda > 0$ and $x \in X$ implies $w(\lambda x) = \lambda^m w(x)$. If X is a cone and D_x is a random variable for each

$x \varepsilon X$, then $\{D_x: x \varepsilon x\}$ is a *conical family* if $\lambda > 0$ and $x \varepsilon X$ implies that D_x and $D_{\lambda x}$ have the same distribution.

It is assumed that the probability space has been extended, if necessary, so that the sample space Ω is a cone.

THEOREM. *Suppose* C *is a cone*, $\{\xi_{s,a}: (s,a) \varepsilon C\}$ *is a conical family*, g *and* J *are homogeneous of degree* m, *and* h *is homogeneous of degree one. Then* f *is homogeneous of degree* m, *i.e.*

$$f(s,\lambda) = \lambda^m f(s/\lambda, 1), \qquad s \varepsilon S,\ \lambda > 0. \tag{2.1}$$

Special cases of the models in [1], [3], [4], [5], [6], and [7] satisfy the conditions of the theorem with m = 1. The contexts are production smoothing, fisheries harvesting, cash management, and inventory control.

Consider a particular model with $\lambda = 1$ and the family of such models as ξ is replaced by $\lambda\xi$. If $f(\cdot,1)$ has been approximated numerically then (2.1) implies that the entire family has been approximated too. Thus the principal virtue of the theorem is its simplification of certain sensitivity analyses in applications of Markov decision processes.

PROOF. Let $\lambda > 0$. From the assumptions, $\lambda^m J[h(z),1] = J[h(\lambda z),\lambda]$, $a \varepsilon A_s \Leftrightarrow (s,a) \varepsilon C \Leftrightarrow (\lambda s, \lambda a) \varepsilon C \Leftrightarrow \lambda a \varepsilon A_{\lambda s}$, and $\xi_{s,a}$ has the same distribution as $\xi_{\lambda s, \lambda a}$. Therefore,

$$\lambda^m f(s,1) = \sup\{E[g(\lambda s, \lambda a, \lambda\xi_{s,a}) + J\{h(\lambda s, \lambda a, \lambda\xi_{s,a}),\lambda\}]: a \varepsilon A_s\}$$

$$= \sup\{E[g(\lambda s, b, \lambda\xi_{\lambda s,b}) + J\{h(\lambda g, b, \lambda\xi_{\lambda s,b}), b\}]: b \varepsilon A_{\lambda s}\}$$

$$= f(\lambda s, \lambda)$$

which is equivalent to (2.1). $\square$

Wijngaard and Boot [7] analyze a version of the production smoothing model in [1]. In that model, $\xi_{s,a}$ is the periodic demand for a product. It is invariant with respect to $(s,a)\varepsilon C$ so $\{\xi_{s,a}: (s,a)\varepsilon C\}$ is trivially a conical family. Wijngaard and Boot restrict the demand distributions in their efforts to prove the following result which is stronger than (2.1). If each ξ is replaced by $\lambda\xi + k$ (where $k\varepsilon\Omega$ and $\lambda\xi + k\varepsilon\Omega$), let $f(\cdot,\lambda, k)$ and $f(\cdot,1,0)$ denote the new and old analogues of (1.3). Then [7] concerns the extent to which $\lambda f(s + k,1,0) = f(s/\lambda,\lambda,k)$ is true. See a paper by Denardo and Rothblum [2] for an entirely different approach to affine dynamic programming.

If $S \subseteq \mathbb{R}$ then (2.1) implies

$$f(s,\lambda) = s^m f(1,\lambda/s), \quad s\varepsilon S(s\neq 0), \quad \lambda > 0, \tag{2.2}$$

which yields computational alternatives to (1.1) and (1.2). In (1.2), for example, let $v(s,\lambda)$ denote $v(s)$ when each $\xi_{s,a}$ is replaced by $\lambda\xi_{s,a}$, and let $V(\lambda) = v(1,\lambda)$. If (2.2) is valid for v, then $v(s,\lambda) = s^m V(\lambda/s)$ so $V(\cdot)$ yields $v(\cdot,1)$. Let

$$G_a(\lambda) = E[g(1,a,\lambda\xi_{1,a})], \qquad H_a(\lambda) = h(1,a,\lambda\xi_{1,a}),$$

and suppose $P\{H_a(\lambda) = 0\} = 0$ for all $\lambda > 0$ and $a\varepsilon A_1$. From (1.2),

$$V(\lambda) = \sup\{G_a(\lambda) + \beta E\{[H_a(\lambda)]^m V[\lambda/H_a(\lambda)]\}: a\varepsilon A_1\}, \quad \lambda > 0. \tag{2.3}$$

The assumption $P\{H_a(\lambda) = 0\} = 0$ is satisfied in [5] and in some cases of [6]. It remains to be seen if there are models for which an approximate numerical solution of (2.3) is an easier task than that of (1.2).

3. Two Applications. Consider a single item discrete-time inventory model in which purchase costs are linear with unit cost c, holding and penalty costs are linear with respective unit costs h and π, demands are backlogged, $\beta < 1$ is the discount factor, and demands $\xi_1, \xi_2, \ldots$ are independent and identically distributed with distribution function $\theta(\cdot)$. Generalizations of this model were studied by Veinott [6].

It follows from [6] that minimization of expected discounted cost leads to minimization of

$$E[\Sigma_{n=1}^{\infty} \beta^{n-1} L_1(a_n)] \tag{3.1}$$

where, for $\lambda > 0$, and $a \in \mathbb{R}$,

$$L_\lambda(a) = c(1-\beta)a + h\int_0^a (a-x)d\theta(x/\lambda) + \pi\int_a^\infty (x-a)d\theta(x/\lambda). \tag{3.2}$$

The minimum of $L_\lambda(\cdot)$ on $\mathbb{R}$ is achieved at

$$a_\lambda^* \triangleq \lambda\theta^{-1}\{[\pi-c(1-\beta)]/(\pi + h)\}, \qquad \lambda > 0 \tag{3.3}$$

and the value of the infimum of (3.1) is $L_1(a_1^*)/(1-\beta)$ if the initial inventory level $s \leq a_1^*$.

Suppose that the problem is altered so each demand ξ_n is replaced by $\lambda\xi_n$, $\lambda > 0$. It follows from [6] that the problem is equivalent to minimization of

(3.1) with $L_\lambda(a_n)$ replacing $L_1(a_n)$. Let $f(s,\lambda)$ denote the value of the minimum if s is the initial inventory level. From [6],

$$f(s,\lambda) = L_\lambda(a_\lambda^*)/(1-\beta), \qquad s \le a_\lambda^*. \tag{3.4}$$

Let K_λ denote the right side of (3.4).

The conditions of the theorem are satisfied with $m = 1$ so

$$f(s,\lambda) = \lambda f(s/\lambda,1). \tag{3.5}$$

From (3.4)

$$f(s,\lambda) = K_\lambda \quad \text{if } s \le a_\lambda^*; \quad \lambda f(s/\lambda,1) = \lambda K_1 \quad \text{if } s/\lambda \le a_1^*.$$

Hence, (3.5) implies

$$a_\lambda^* = \lambda a_1^* \text{ and } K_\lambda = \lambda K_1. \tag{3.6}$$

Heretofore, (3.6) has been known only for the several specific classes of distribution functions $\theta(\cdot)$ for which a_λ^* and $L_\lambda(a_\lambda^*)$ can be written explicitly.

Consider a fishery with q species ("species" may include different age-classes of the same species) in which the objective is to maximize the expected discounted revenue from all species caught. If a_n is the vector of population levels (in units of biomass) at the end of the n-*th* fishing season, then assume that the vector of population levels (in units of biomass) at the start of the

$n+1$-*st* season is distributed as $h(a_n,\xi)$. Let ρ denote the vector of prices (per unit biomass) for the species. This model is related to [3] and [4] and leads to the dynamic program

$$f(s,\lambda) = \sup\{\rho\cdot(s-a) + \beta E\{f[h(a,\xi),\ \lambda]\}:\ 0 \le a \le s\} \tag{3.7}$$

in which $\lambda > 0$ and $s \in \mathbb{R}^q$ with $s \geq 0$.

Suppose that the expectation in (3.7) exists for all a and that (3.7) is finite for all s and λ. Suppose too that a_λ^* maximizes $-\rho\cdot a + \beta E\{f[h(a,\xi),\ \lambda]\}$ on the nonnegative orthant of $\mathbb{R}^q$. Then

$$f(s,\lambda) = K_\lambda \qquad \text{if } a_\lambda^* \le s, \tag{3.8}$$

$$K_\lambda \triangleq \rho\cdot(s-a_\lambda^*) + \beta E\{f[h(a_\lambda^*,\xi),\ \lambda]\}.$$

If h is homogeneous of degree one, then the theorem's conditions are satisfied so, from the argument yielding (3.6),

$$a_\lambda^* = \lambda a_1^* \ \text{ and } K_\lambda = \lambda K_1. \tag{3.9}$$

For $x = (x_1, x_2, \ldots, x_q) \in \mathbb{R}^q$, let $(x)^+$ denote the q-vector whose i-*th* component is $\max\{x_i, 0\}$. An example of a homogeneous h which suffices for some fisheries is

$$h(a,\xi) = (Ma + \xi)^+$$

where M is a $q \times q$ real matrix and ξ is a q-vector valued random variable.

Acknowledgements. This research was partially supported by National Science Foundation Grant SOC78-T5730.

REFERENCES

1. Beckmann, M. J. (1961). Production Smoothing and Inventory Control. Operations Research 9 456-467.

2. Denardo, E. V. and Rothblum, U. G. (1978). Affine Dynamic Programming, in Dynamic Programming and its Applications ed. by M. L. Puterman Academic Press, New York.

3. Mendelssohn, R. (1978). Optimal Harvesting Strategies for Stochastic Single Species, Multiage Class Models. Mathematical Biosciences 41 159-174.

4. Mendelssohn, R. and Sobel, M. J. (1980). Capital Accumulation and the Optimization of Renewable Resource Models. J. of Economic Theory 23 to appear.

5. Porteus, E. L. (1972). Equivalent Formulations of the Stochastic Cash Balance Problem. Management Science 19 250-253.

6. Veinott, A. F. Jr. (1965). Optimal Policy for a Multi-Product, Dynamic, Nonstationary Inventory Problem. Management Science 12 206-222.

7. Wijngaard, J. and Boot, J. G. C. (1980). A Note on the Cost of Uncertainty in a Production Smoothing Problem. Dept. of Industrial Engineering, Eindhoven University (Netherlands) report BDK/ORS/80/04.

Stochastic Fishery Games with Myopic Equilibria

by

Matthew J. Sobel
Georgia Institute of Technology

MS-80-4

STOCHASTIC FISHERY GAMES WITH MYOPIC EQUILIBRIA*

by

Matthew J. Sobel**
Georgia Institute of Technology

Abstract

There are strong reasons for fishery managers to consider interactions between fish species and between fishing boats, processors, and consumers. Sequential stochastic games are natural models for these phenomena and this paper formulates several such models. In the case of linear returns to each "player", simple conditions are given that ensure existence of a *myopic* equilibrium point. Myopic equilibria can be computed easily. A Markov decision process is a stochastic game with only one player so some sequential control models of fisheries possess myopic optima. Several of the papers presented at this Conference contain models whose optima are myopic.

*Presented at the University of Chicago, Georgia Institute of Technology, and Yale University in 1976 and 1977 and at a Conference on the Economics of Renewable Resource Management, Brown University, October 1979.

**Partially supported by NSF grant SOC 78-05770.

STOCHASTIC FISHERY GAMES WITH MYOPIC EQUILIBRIA

by

Matthew J. Sobel
Georgia Institute of Technology

1. Introduction

Most implemented fishery models use a single pooled age class. If the model concerns more than one species, then the species are pooled too. The reasons given to justify such highly aggregated models are mathematical complexity, numerical complexity, and sparse data. Many scientists and regulatory agencies realize that less highly aggregated models are desireable planning tools.

Different species of fish may prey on one another or compete for common prey (Figure 1). For example, in New England, cod, pollock, and siler hake feed on herring; mackerel and silver hake feed on each other; and cod feed on mackerel. The age-class structure of a single species often exhibits marked fluctuations as time passes. These fluctuations should affect harvesting policies so there are biological reasons to build multi-species multi-age-class models.

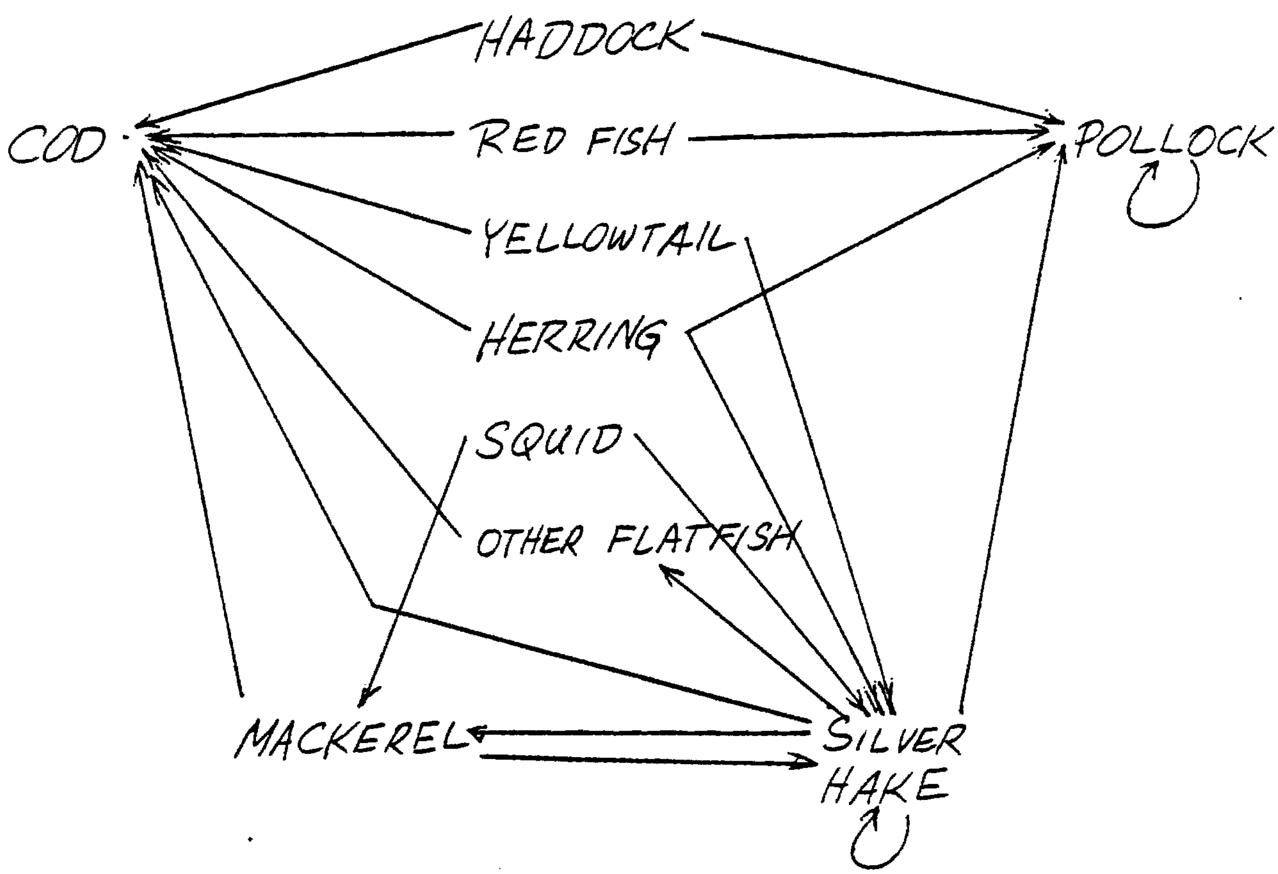
The Fishery Conservation and Management Act of 1976 (P.L. 94-265) mandates a regional approach to fishery management, Therefore, the biological interactions mentioned above preclude disaggregation of the management task into independent tasks of managing each species separately. The Act creates Regional Councils so now there are instititional and legal reasons to build multi-species models.

Some species of fish are caught commercially together with other species. This "bycatch" phenomenon is a technological reason to build multi-species models.

These technological and biological considerations provide economic incentives to coordinate the management of several species. From Figure 1, in New England we should not stipulate the mackerel catch without considering the effects on cod and silver hake.

Figure 1

NEW ENGLAND FISHERY



OTHER FINFISH ARE FOOD FOR ALL; THEY PREY ON ALL EXCEPT COD & POLLOCK.

OTHER INVERTEBRATES ARE FOOD FOR ALL.

X ⟶ Y DENOTES Y FEEDS ON X.

The stochastic game in §2 includes a general class of species and age-class interactions in the underlying population dynamics. The model is presented in §2 and critiqued in §4. The special case of linear returns leads to a simplification which is presented formally in §3. The potential applications to fisheries are discussed in §5. In this case, the stochastic game has a "myopic" equilibrium point which can be computed relatively easily.

There is a large and growing literature on the 1976 Act. Other writers have noted the multiplicity of objectives and "players" in a regional fishery (cf. Mendelssohn [6]). Some writers have analyzed the international "game" in conflicts over fisheries resources (cf. Levhari and Mirman [5]). This paper proposes that *sequential* game models be used to analyze dynamic interactions. Many others (cf. Clark [2]) have advocated the use of sequential models to analyze the optimal operation of renewable resources where "optimal" refers to a single perspective (firm, agency, country, etc.).

2. A Stochastic Fishery Game

Let I be the set of relevant species. The species identities are assumed to be equivalent to the various sub-industries that fish for the species. In year t, t=1,2, ..., the *state* is the vector s_t. The i-*th* component of s_t is s_t^i which is the biomass of species i at the beginning of the fishing season. Let S denote the set of all possible states. If $s \in S$ then s^i denotes the i-*th* component of s. This notation is consistent with a more general model that has several age-classes for each species. Then s_t^i and s^i become vectors with as many components as there are age-classes for species i. The same remark applies to the following notation.

Let a_t^i denote the biomass of species i at the end of the fishing season in year t, i.e. its "escapement", and let $a_t = (a_t^i, i \in I)$. The set of possible values of a_t depends, generally, on s_t. For example, $0 \leq a_t^i \leq s_t^i$ unless species i reproduces extraordinarily quickly. If the constraint is $0 \le a_t^i \le s_t^i$, we might interpret $s_t^i - a_t^i$ as the size of the catch of species i in year t. This interpretation is reasonable only if natural mortality and recruitment are negligible during the fishing season. Generally, let A_s^i denote the set of feasible values of a_t^i if $s_t = s$. Then $C_s = \bigtimes_{i\varepsilon I} A_s^i$ is the set of feasible values of a_t if $s_t = s$ and $a_t \varepsilon C_{s_t} \Leftrightarrow a_t^i \varepsilon A_{s_t}^i$ for all $i\varepsilon I$. Let $A = \bigcup_{s\varepsilon S} C_s$. If $a\varepsilon A$ then a^i denotes the i-*th* component of a. For any t, $W = \{(s,a)\colon a\varepsilon C_s,\ s\varepsilon S\}$ denotes the set of feasible (s_t,a_t) pairs. If $A_s^i = [0,s^i]$ for each i, then $W = \{(s,a)\colon 0 \le a \le s,\ s\varepsilon S\}$. For computational and expository convenience, *assume that* W *is a set with only finitely many elements.*

The population dynamics are specified by transition probabilities. Assume that (i) the dynamics are Markovian so s_{t+1} has a distribution that depends on $s_1,a_1,\ldots,s_t,a_t$ only through s_t and a_t; (ii) the distribution of s_{t+1} does not depend on s_t. Assumption (i) is made throughout the stochastic game literature and is reasonable for a fishery. The biomasses in year t+1 depend on the biomasses and catches in year t. Assumption (ii) is a characteristic of regional fisheries that significantly simplifies the analysis of their dynamics.

Assumptions (i) and (ii) together imply that there are numbers $p_j(a)$, $j\varepsilon S$ and $a\varepsilon A$, such that

$$(1)\qquad \begin{cases} p_j(a) = P\{s_{t+1} = j \mid a_t = a\} \\ \qquad\quad = P\{s_{t+1} = j \mid s_1,a_1,\ldots,s_t,a_t = a\} \end{cases}$$

for all t. Therefore, the biomasses at the start of the t+1-*st* fishing season are assumed to depend only on the biomasses at the end of the t-*th* season (and, possibly, on exogenous uncertainty). In fact, (1) is implicit in most theoretical and empirical fisheries research. It is a stochastic generalization of the "metered model" in [1].

Let α_i, $0 \le \alpha_i < 1$, denote the single period discount factor for the sub-industry which harvests species i. Let $r_i(s,a)$ denote sub-industry i's (expected) profit in any year t where $s_t=s$ and $a_t=a$. For example, if (i) sub-industry i's profit depends only on the catch of species i, (ii) $s^i - a^i$ is the catch of species i, and (iii) if its profit is proportional to the size of the catch, then

$$(2) \qquad r_i(s,a) = b_i \cdot (s^i - a^i)$$

where b_i is the unit profit. Let

$$(3) \qquad V_i = \Sigma_{t=1}^{\infty} \alpha_i^{t-1} \, r_i(s_t, a_t)$$

denote sub-industry i's sum of discounted profits.

Let π^i denote the set of all nonanticipative decision rules (possibly history dependent and randomized) which player i could use to choose the sequence of actions $a_1^i, a_2^i, \ldots$. Let $\pi = \underset{i\varepsilon I}{\times} \pi^i$ be the product set of all players' possible rules and, for $\rho\varepsilon\pi$, write $\rho = (\rho^i, \rho^{-i})$ where $\rho^{-i} \varepsilon \prod_{j\neq i} \pi^j$.

For $\rho\varepsilon\pi$, the expectation $v_i(\rho|s) = E_\rho[V_i | s_1 = s]$ exists for all $i\varepsilon I$, $s\varepsilon S$, and $\rho\varepsilon\pi$.

DEFINITION. *A policy* $\delta\varepsilon\pi$ *is an* EQUILIBRIUM POINT RELATIVE TO $X \subseteq S$ *iff*

$$(4) \qquad v_i(\delta|s) = \sup\{v_i(\rho\ \delta^{-i})|s) : \ \rho\varepsilon\pi^i\}, \qquad s\varepsilon X, \qquad i\varepsilon I.$$

Thus an equilibrium point relative to X is noncollusively optimal for every player for each initial state in X.

There is a stationary policy which is an equilibrium point relative to S [11,4,9,13]. However, even a two-person zero sum stochastic game with $|W| < \infty$ and rational data may possess an irrational value. Therefore, a finitely convergent algorithm is not achievable in general.

3. Myopic Equilibria

Assumption (1) and two others below lead to a relatively easily computable equilibrium point relative to a proper subset of S. Suppose

(5) $$r_i(s,a) = K_i(a) + L_i(s), \quad i\varepsilon I, \quad (s,a)\varepsilon W.$$

The case which is pertinent for fisheries is (2) where $K_i(a) = b_i a^i$ and $L_i(s) = b_i s^i$. Under assumptions (1) and (5), let

(6) $$G_i(a) = K_i(a) + \alpha_i \Sigma_{s\varepsilon S}\, p_s(a) L_i(s), \quad a\varepsilon A, \quad i\varepsilon I.$$

When (5) is caused by (2),

(7) $$G_i(a) = -b_i a^i + \alpha_i b_i \Sigma_{s\varepsilon S}\, p_s(a) s^i$$

Let $S(a) = \{s\colon\ s\varepsilon S,\ a\varepsilon A_s\}$ denote the states in which action $a\varepsilon A$ is feasible. Let Γ denote the following noncooperative static (Nash) game among the players in I. Player $i\varepsilon I$ has available the set $A^i = \bigcup_{s\varepsilon S} A^i_s$ of alternative actions and $G_i(a)$ is player i's payoff if the actions are $a = (a^i,\ i\varepsilon I)$. Since $|A^i| < \infty$ for each i and $|I| < \infty$, there is an equilibrium point of Γ. First, suppose that a_* is an unrandomized equilibrium point of Γ:

(8) *There exists* $a_*\varepsilon A$ *which is an equilibrium point of* Γ.

THEOREM. *Assumptions* (1), (5), (8), *and*

(9) $$\Sigma_{j\varepsilon S(a_*)}\, p_j(a_*) = 1$$

imply $a_t = a_*$ *for all* $t=1,2,\ \ldots$ *is an equilibrium point relative to* $S(a_*)$.

There is a more general version of this result where the equilibrium point may represent randomized strategies. The more general version is the Corollary below. The significance of the Theorem and Corollary is that there are several finitely convergent algorithms to compute an equilibrium point of a game such as Γ (a Nash game). The algorithms are cited in [14] which proves the Theorem and Corollary. A stochastic game that satisfies the conditions of the Theorem or Corollary is said to have a *myopic equilibrium* point. *Ad infinitum* repetition of the equilibrium point of a static game comprises an equilibrium point of the dynamic game.

For the Corollary, let D_s^i denote the set of all probability vectors on A_s^i and $D = \bigtimes_{i\in I} \bigcup_{s\in S} D_s^i$ which is the randomization version of $A = \bigtimes_{i\in I} \bigcup_{s\in S} A_s^i$.

(10) *Let* $\underset{\sim}{d}\in D$ *denote an equilibrium point of* Γ

which necessarily exists because $|A| < \infty$. In this notation,

$$\underset{\sim}{d}(a) = \prod_{i\in I} P\{\text{in } \Gamma \text{ player } i \text{ chooses action } a^i \in A^i\}.$$

Let

$$A' = \{a: \ a\in A, \ \underset{\sim}{d}(a) > 0\}$$

$$S' = \{s: \ s\in S, \ a\in A_s \text{ for all } a\in A'\} = \bigcup_{a\in A'} S(a).$$

COROLLARY. *Assumptions* (1), (5), *and*

(11) $\qquad \Sigma_{j\varepsilon S'} \, p_j(a) = 1 \qquad$ *for all* $a\varepsilon A'$

imply $P\{a_t = a\} = \underset{\sim}{d}(a)$ *for all* $t=1,2,\ldots$ *is an equilibrium point relative to* S'.

4. Critique of Assumptions

The applicability of the Theorem and Corollary depends on the extent to which the assumptions are reasonable in fisheries. Assumption (2), $r_i(s,a) = b_i(s^i-a^i)$, and its consequences are discussed in §4. It is implicit in (2) that s^i-a^i is interpreted as the catch of species i. We have already observed that this interpretation is reasonable if natural mortality and recruitment are not too massive during the fishing season.

The stochastic game model in §2 uses a_t^i, the biomass of species i in the fishery at the end of the t-*th* fishing season, as the generic decision variable. In practice, often the estimates of such biomasses are crude. Actually one has only sample data from which to infer posterior distributions of s_t and a_t and these distributions may exhibit significant variation. Is catch size a more appropriate decision variable because it can be measured accurately? It is a Hobson's choice.

Let $\xi(a)$ denote a random variable with the distribution given in (1):

$$p_j(a) = P\{\xi(a) = j\}, \qquad j\varepsilon S,\ a\varepsilon A.$$

Then (1) asserts $s_{t+1} \sim \xi(a_t)$ where $X \sim Y$ indicates that random variables X and Y have the same distribution. The preceding paragraph admits that we have only sample data concerning s_t. Suppose X_t are the sample data and let z_t^i denote the catch size $s_t^i - a_t^i$. Then $s_{t+1} \sim \xi(s_t - z_t)$ but we have to settle for

(12) $$s_{t+1} \sim E[\xi(\underset{\sim}{s} - z_t)|X_t]$$

where the expectation is taken with respect to the conditional distribution of s_t given X_t. In words, even if the decision variable is catch size rather than residual biomass, (12) exhibits the same posterior distribution for s_{t+1}.

Assumption (9) has the following interpretation in a fishery model where $s_t^i - a_t^i$ indicates catch size:

(13) $$P\{s_{t+1}^i \geq a_*^i\} = 1, \quad i\varepsilon I,\ t = 1,2, \ldots .$$

More generally, in the notation of the Corollary let

$$a_o^i = \max\{a^i: \ a = (a^i,a^{-i})\varepsilon A'\}$$

and $a_o = (a_o^i,\ i\varepsilon I)$. In the notation of the preceeding paragraph, assumption (11) is equivalent to

(14) $$P_{\underset{\sim}{d}}\{\xi(\underset{\sim}{a}) \geq a_o\} = 1$$

where $\underset{\sim}{a}$ is the random action that will be taken using $\underset{\sim}{d}$.

In a fishery model, $\xi(\cdot)$ is usually stochastically nondecreasing (this ignores overpopulation, i.e. "overcompensation"), so

$$(15) \qquad P\{\xi(a') \leq b\} \geq P\{\xi(a'') \leq b\}, \qquad a' \leq a'',$$

for every nonnegative I-vector b. Let

$$a_+^i = \min\{a^i: \ a = (a^i, a^{-i}) \varepsilon A'\}$$

and $a_+ = (a_+^i, \ i \varepsilon I)$. Then (11) and (15) require

$$(16) \qquad P\{\xi(a_+) \geq a_o\} = 1.$$

The general sense of (9), (11), (13), (14), or (16) is that, with probability one, recruitment minus natural mortality exceeds catch size. If any component of a_* (or a_+) is very low then the assumption is likely to be unreasonable ("overfishing.").

It is implicit in §2 that a separate sub-industry harvests each major commercial species. Without such an assumption, consider a dynamic model of equilibrium between sub-industries. Such a model would still be a stochastic game and an equilibrium point would necessarily exist. But computation and characterization would be more difficult (or impossible) because the Theorem and Corollary would be invalid.

The identification of sub-industries with species may be a reasonable first approximation in some contexts. Only empirical results can support a conclusion regarding whether or not the courseness of the approximation is an acceptable price to pay for relatively easy computations.

5. Fisheries Applications of Myopic Equilibria

Linear returns (2) leads to (7) which may be written

$$(17) \qquad G_i(a) = b_i \Sigma_{s \varepsilon S}\, p_s(a)(\alpha_i s^i - a^i).$$

It is apparent that the set of equilibrium points of Γ does not depend on $b = (b_i, i\varepsilon I)$. Also, $G_i(a)$ is proportional to a weighted average of the difference between the discounted biomass at the start of next year's season and the biomass at the end of this year's season.

The myopic equilibrium asserted in the Theorem is closely related to a result in [7]. The model there is a (single player) version of the present one where there is exactly one species. Then linear returns (and some technical assumptions) yields a myopic optimum without the restriction (9), i.e. without requiring feasibility with probability one.

The decision rule in the Theorem and Corollary is part of a *stationary* policy. A policy $\rho\varepsilon\pi$ is said to be stationary if there is an element $\delta\varepsilon D$ such that $\delta(s_t) = P\{a_t = a\}$ for all $t=1,2, \ldots$. Thus the actions are not necessarily the same for all t but the rule that determines the actions *is* the same. What are the ergodic consequences of using a stationary policy in a stochastic game? Section 6 in [7] contributes to the literature (see references in [7]) which answers this question for single species models. The primary tool in [§6 of 7] is the general theory of discrete-time Markov processes. The same theory was used by Sanghvi [10] to obtain some ergodic consequences of stationary policies in stochastic games. He focuses on probabilities of absorbtion.

In most fisheries models, the absorbing states connote "extinction" so Sanghvi's results may be useful here.

Consider a variant of the model in §2 which stresses aggregate regional benefits. Instead of a game-like perspective, suppose that the objective is to maximize the expected discounted aggregate benefits:

$$\text{maximize } E[\Sigma_{t=1}^{\infty}\Sigma_{i\varepsilon I}\alpha_i^{t-1}\ r_i(s_t,a_t)].$$

Let

$$(18) \quad G(a) = \Sigma_{i\varepsilon I}\ G_i(a) = \Sigma_{s\varepsilon S}\ p_s(a)\Sigma_{i\varepsilon I}b_i(\alpha_i s^i - a^i), \quad a\varepsilon A,$$

and let a_* denote an element of A that maximizes $G(\cdot)$ on A. Such an element exists because A is a finite set. Suppose $C_s = [0,s]$ for $s\varepsilon S$. As a consequence of the Theorem, assumptions (1), (2), and (9) imply $a_t = a_*$ for all t=1,2, ... is optimal with respect to all initial states $s \geq a_*$. Such a policy is called a *myopic optimum.* This result can be easily extended to a more general model where (i) each species has an age structure, and (ii) sub-industries are not equivalent to species.

When $|I| = 1$ so species (and age-classes) are pooled, a myopic optimum in (18) is merely the familiar discounted version of maximum sustained yield. Suppress subscript i in (18) because $I = \{1\}$ and let $\mu(a)$ denote $\Sigma_{s\varepsilon S}p_s(a)s$. Then (18) is the familiar relationship

$$G(a) = b[\alpha\mu(a) - a].$$

If A were an (open) interval and $\mu(\cdot)$ were continuously differentiable, then a necessary condition for a_* would be $\alpha\mu'(a_*) - 1 = 0$ or $\mu'(a_*) = \alpha^{-1}$. This stochastic version of the usual maximum sustained yield condition occurs in the model (with K=0) that Daniel F. Spulber [15] presented at this Conference.

Another example of a myopic optimum occurs in a stochastic version of the model that Lee G. Anderson presented at this Conference [1]. His model concerns fleet size and allocation of fishing effort between two species. The following model addresses only allocation of effort. For each n, s_n, a_n, and $\xi(a_n)$ are two-vectors with a component for each species. In fact, the objective in [1] leads to (18) with $|I| = 2$. The distinghishing feature in [1] is an upper bound on a linear combination of the catches of the two species:

$$w_1 \cdot (s_n^1 - a_n^1) + w_2 \cdot (s_n^2 - a_n^2) \leq u \tag{19}$$

where u denotes the maximum feasible fishing effort. Suppose that $a_* = (a_*^1, a_*^2)$ maximizes (18). Then (19) implies that (9) is equivalent to

$$P\{\xi(a_*) \geq a_*,\ w \cdot (\xi[a_*] - a_*) \leq u\} = 1$$

where $w = (w_1, w_2)$.

Colin W. Clark's paper [3] at this Conference includes a duopoly model that does *not* have a myopic equilibrium point. His model has two players; player 1 is the fishing fleet and player 2 is the processor. Let a^1 denote the escapement chosen by the fleet and let a^2 denote the unit price of fish that the processor will pay the fleet. Then the single period rewards are

$$r_1(s,a) = (a^2 - c)(s - a^1), \qquad r_2(s,a) = (\rho - a^2)(s - a^1)$$

where s is the biomass at the beginning of the season, $a = (a^1, a^2)$, and ρ is the wholesale price (less processing cost) that the processor receives for processed fish. Therefore,

$$r_1(s,a) = a^2 s - cs + ca^1 - a^1 a^2$$

$$r_2(s,a) = -a^2 s + \rho s - \rho a^1 + a^2 a^1$$

whose first terms lack the additive decomposition feature in (5).

REFERENCES

1. Anderson, L.G., "The Economics of Multipurpose Fleet Behavior," presented at this Conference (1979).

2. Clark, C.W., Mathematical Bioeconomics, Wiley, New York (1976).

3. ___________, "Towards a Predictive Model for the Economic Regulation of Commercial Fisheries," presented at this Conference (1979).

4. Fink, A.M., "Equilibrium in a Stochastic n-Person Game," J. Sci. Hiroshima Univ. Ser. A-I 28 (1964), 89-93.

5. Levhari, D. and L. Mirman, "The Great Fish War: An Example Using a Cournot-Nash Equilibrium," Bell J., to appear.

6. Mendelssohn, R., "Harvesting Policies and the Fishery Conservation and Management Act of 1976," unpublished manuscript (1977).

7. ______________, and M.J. Sobel, "Capital Accumulation and the Optimization of Renewable Resource Models," Journal of Economic Theory, to appear.

8. Parthesarthy, T. and M. Stern, "On Markov Games," Lecture Notes in Pure and Applied Mathematics, Vol. 30 (1977) Marcel Dekker Publishing Company, pp. 1-46.

9. Rogers, P. D., "Nonzero-sum stochastic games. Report ORC 69-8, Operations Research Center, University of California, Berkeley, 1969.

10. Sanghvi, A. P., "Sequential Games as Stochastic Processes," Stochastic Processes and their Applications, 6 (1978), 323-336.

11. Shapley, L., "Stochastic Games," Proceedings of the National Academy of Sciences, USA, Vol. 39 (1953), pp. 1095-1100.

12. Shubik, M. and M. J. Sobel, "Stochastic Games, Oligopoly Theory, and Competitive Resource Allocation," to appear in Dynamic Optimization and Mathematical Economics edited by P. T. Liu, Vol. 19 in the series 'Mathematical Concepts and Methods in Science and Engineering," Plenum Press, New York, L970.

13. Sobel, M. J., "Noncooperative Stochastic Games," Annals of Mathematical Statistics, Vol. 42 (1971), pp. 1930-1935.

14. _____________ "Myopic Equilibrium Points in Stochastic Games," unpublished manuscript, 1979.

15. Spulber, D.F., "Adaptive Harvesting of a Renewable Resource and Stable Equilibrium," presented at this Conference (1979).

16. Whitt, W., "Representation and Approximation of Noncooperative Stochastic Games," J. SIAM Control, to appear.

MANAGEMENT SCIENCE WORKING PAPERS

MS-78-6 : Bernell K. Stone, "The Design of a Receivable Collection System: Heuristic Procedures."

MS-78-7 : Robert P. Kertz and David C. Nachman, "Persistently Optimal Plans for Non-Stationary Dynamic Programming: The Topology of Weak Convergence Case." Revision of MS-77-1.

MS-78-8 : George E. Monahan, "Optimal Stopping in a Partially Ovservable Markov Process with Costly Information."

MS-78-9 : Bernell K. Stone, "Lock Box Selection and Collection System Design: Objective Function Validity."

MS-78-10: Bernell K. Stone, "Break-Even Receivable Size and the Allocation of Receivables to Lock Boxes."

MS-79-1 : Bernell K. Stone, "Zero-Balance Banking and Collection System Design In a Divisionalized Firm."

MS-79-2 : Bernell K. Stone, "Design of a Receivable Collection System: Sequential Building Heuristics."

MS-79-3 : Bernell K. Stone, "Daily Cash Forecasting: A Multiplicative Model for Measuring Cash Flow Cycles."

MS-79-4: George E. Monahan, "Optimal Stopping in a Partially Observable, Binary-Valued Markov Chain with Costly Perfect Information." Revised April, 1979.

MS-79-5: Matthew Sobel and Martin Shubik, "Stochastic Games, Oligopoly Theory and Competitive Resource Allocation."

MS-79-6: Bernell K. Stone, Ned Hill, "Cash Transfer Scheduling: The Key to Efficient Cas Con."

MS-79-7: Bernell K. Stone, Ned Hill, "The Design of a Cash Concentration System."

MS-79-8: Matthew Sobel, "Simple Inequalities for Multiserver Queues."

MS-79-9: Matthew J. Sobel, "Capital Accumulation and the Optimization of Renewable Resource Models." With Roy Mendelssohn.

MS-79-10: Robert G. Jeroslow, "A Disjunctive Penalty Method For Some Strong Nonconvexities."

IS-79-11: George E. Monahan, "Optimal Stopping in a Partially Observable Markov Process with Costly Information." Revised June, 1979.

S-79-12: George E. Monahan, "Optimal Stopping in a Partially Observable Markov Process with Costly Information." Revised October 1979.

MANAGEMENT SCIENCE WORKING PAPERS

MS-29-13: Robert G. Jeroslow, "The Limiting Lagrangean." With R. J. Duffin

MS-79-14: Robert G. Jeroslow, "A Limiting INFISUP Theorem." With C. E. Blair and R. J. Duffin.

MS-79-15: Robert G. Jeroslow, "An Exact Penalty Method for Mixed-Integer Programs." With C. E. Blair.

MS-79-16: Bernell K. Stone and Tom W. Miller, "The Use of Constrained Regression Estimates in Daily Cash Flow Forecasting."

MS-79-17: Bernell K. Stone and Tom W. Miller, "Forecasting Short-Term Cash Flows With Constrained Growth Models."

MS-79-18 Bernell K. Stone and Tom W. Miller, "Daily Cash Forecasting With Dummy-Variable Regression Using Multiplicative and Mixed-Effects Models for Measuring Cash Flow Cycles."

MS-80-1: Bernell K. Stone and Ned C. Hill, "Cash Transfer Scheduling for Efficient Cash Concentration."

MS-80-2: Bernell K. Stone and Ned C. Hill, "The Evaluation of Alternative Cash Transfer Mechanisms and Methods."

MS-80-3: Mathew J. Sobel, "Ordinal Sequential Games."

MS-80-4: Matthew J. Sobel, "Stochastic Fishery Games With Myopic Equilibria."

Chapter VIII

SEQUENTIAL GAMES

A sequential game is a multi-person decision process in which each participant makes a sequence of decisions. The participants' sequences of decisions influence the evolution of the process and affect the time streams of rewards to the participants. A sequential game is also called a *stochastic game* and a *Markov game*.

Sequential game models have been constructed of diverse phenomena in management science, biology, economics, psychology, and military affairs. Some of the phenomena are arms control and disarmament, advertising decisions of competing firms, pricing and production decisions of competing firms, interactions of biological species, harvesting decisions in a fishery, the entry and exit of firms to and from an industry, pursuit-evasion tactics for opposing submarines, duels between opposing aircraft, and various paradigms in experimental social psychology.

The dynamic games in this chapter are discrete in time.[1] There is a largely separate literature on continuous-time sequential games called *differential games*. The Bibliographic Guide at the end of the chapter includes several general references on differential games.

A sequential game is a natural generalization both of Markov decision processes (MDP's) and "static" game theory.[2] Therefore, this chapter leans heavily on earlier chapters.

---

[1] They encompass games based on semi-Markov processes in the sense that §V.1 demonstrates that a semi-Markov decision process is equivalent to an MDP.

[2] We are rather unfair in our perjorative use of "static". First, the "static" normal form of games encompasses dynamic models. Second, the theory of sequential games originated in the late 1940's so it has, by now, attained classical status.

Chapter IV influences the entire chapter, Chapter III influences §4 and §6, and §V.2 is used in §3.

Section	Topic
1	The Model
2	Concepts of Solution
3*	Existence of Solutions
4	Myopic Equilibrium Points
5*	Competitive Advertising Decisions
6*	Dynamic Oligopoly
7*	Ordinal Sequential Games
8*	Algorithms

1. The Model

The canonical elements of a sequential game model include nonempty sets $\mathcal{Q}$ of *players*, S of *states*, and A_s^q of alternative *actions* for each $q \varepsilon Q$ and $s \varepsilon S$. The other canonical elements are a *transition function* and *single stage reward functions*. The general idea is that (i) each player must take a sequence of decisions, and (ii) the state of affairs when the players are about to take their n-*th* decisions (for each n) is adequately summarized by some state $s \varepsilon S$.

The "adequately" in (ii) has three qualifications. First, the constraining effects of the past history on player q's set of feasible alternative actions is completely specified by s in A_s^q. Second, the subsequent state of affairs is conditionally independent of the past history given the current state and current action. Third, the immediate reward, possibly an r.v., also is conditionally independent of the past history given the current state and current action. These are Markovian assumptions.

The sample path (or outcome) of a sequential game specifies the successive states and actions. Let a_n^q denote the action taken by player q in period n, let $\underset{\sim}{a}_n = (a_n^q,\ q \varepsilon \mathcal{Q})$ be the vector of all players' actions in period n, and let s_n denote the state at the beginning of period n. In order to be consistent with our MDP notation, let

$$A_s = \times_{q \varepsilon \mathcal{Q}} A_s^q$$

$$\mathcal{S} = \{(s,a)\colon a \varepsilon A_s, \quad s \varepsilon S\}$$

Hence, A_s is the set of feasible action vectors (actions of all players) in period n if the state is $s_n = s$.

Definitions and Assumptions

Only in this subsection, we use boldface type to denote random quantities and ordinary typeface to denote the values taken by random quantities.[1]

DEFINITION 1. *The* history *up to the time at which the n-th action is taken is*

(1) $$H_n \triangleq (s_1, a_1, s_2, a_2, \ldots, s_{n-1}, a_{n-1}, s_n).$$

As with an MDP, at time $j < n$, all the players know H_j but $\underset{\sim}{s}_n$ and $\underset{\sim}{a}_n$ are not known. They are r.v.'s because they depend on how the game evolves during $j+1, j+2, \ldots, n$.

ASSUMPTION 1. For any $J \subset S$,

(2) $$P\{\underset{\sim}{s}_{n+1} \varepsilon J \mid \underset{\sim}{H}_n, \underset{\sim}{a}_n\} = P\{\underset{\sim}{s}_{n+1} \varepsilon J \mid \underset{\sim}{s}_n, \underset{\sim}{a}_n\}.$$

This Markovian assumption implies existence of a *transition function* $p(\cdot|\cdot,\cdot)$ defined as follows.

DEFINITION 2. *The* transition function $p(\cdot|\cdot,\cdot)$ *satisfies*

$$p(J|s,a) = P\{\underset{\sim}{s}_{n+1} \varepsilon J \mid \underset{\sim}{s}_n = s, \underset{\sim}{a}_n = a\}, \quad n \varepsilon I_+, \quad (s,a) \varepsilon \mathcal{G}.$$

When S is a discrete set, we write $p_{sj}(a)$ for $p(\{j\}|s,a)$.

Here is the Markovian assumption concerning the reward $\underset{\sim}{X}_{nq}$ received by player q in period n.

ASSUMPTION 2. *For each* $n \varepsilon I_+$ *and* $q \varepsilon Q$, $E(\underset{\sim}{X}_{nq} | \underset{\sim}{H}_n; \underset{\sim}{a}_n) = E(\underset{\sim}{X}_{nq} | \underset{\sim}{s}_n, \underset{\sim}{a}_n)$.

DEFINITION 3. *The* single stage reward function *is*

---

[1] Elsewhere in the book (except for Sections 1 and 2 of Chapter IV), boldface type denotes a vector.

$$r_q(s,a) = E(\underset{\sim}{X}_{nq} | \underset{\sim}{s}_n = s, \underset{\sim}{a}_n = a), \qquad (s,a) \varepsilon \mathcal{G}, \qquad q \varepsilon \mathcal{Q}.$$

The extant sequential game theory is based on the assumption that all the players have the same planning horizon. Let N denote that planning horizon (possibly infinity). We call N the *duration* of the game. If $N < \infty$, let $L_q(s)$ denote the salvage value received by player q if the ultimate state is $s_{N+1} = s$.

DEFINITION 4. *The* duration *of the game N is a positive integer or infinity. The* salvage value function *is*

$$L_q(s) = E(X_{N+1,q} | s_{N+1} = s), \qquad s \varepsilon S, \qquad q \varepsilon \mathcal{Q}.$$

DEFINITION 5. *A* sequential game (SG) *is a model which consists of the nonempty sets* $\mathcal{Q}$, S, *and* A_s^q, $s \varepsilon S$ *and* $q \varepsilon \mathcal{Q}$, *the transition function, the single stage reward function, the duration* N, *and if* $N < \infty$ *the salvage value function, and which satisfies Assumptions* 1 *and* 2.

Let Q denote the number of players, i.e. number of elements in the set $\mathcal{Q}$.

DEFINITION 6. *An SG is* finite *if* $\mathcal{G}$ *is a finite set.*

It follows from the definition that a finite SG has only finitely many players, finitely many states, and each player in each state has only finitely many alternative actions.

DEFINITION 6. *An SG is* finite *if* $\mathcal{G}$ *is a finite set.*

Let β_q be player q's discount factor, $0 \leq \beta_q \leq 1$. Most solution concepts for SG's (as is true for MDP's) concern expected values of the following r.v.'s:

(3a) $$V^q(N) = \Sigma_{n=1}^{N} \beta_q^{n-1} r_q(s_n, a_n) + \beta_q^N L_q(s_{N+1}) \qquad (N < \infty),$$

(3b) $$V^q = \Sigma_{n=1}^{\infty} \beta_q^{n-1} r_q(s_n, a_n) \qquad (\beta_q < 1),$$

and

(3c) $$G^q = \lim\inf_{M \to \infty} \Sigma_{n=1}^{M} r(s_n, a_n)/M.$$

Of course, these are the sums of discounted rewards for finite and infinite durations, and the average reward per period.

In Sections 2 and 3 of Chapter IV we discuss the notion of a policy in an MDP. Briefly a policy in an MDP is a nonanticipative and deterministic contingency plan for making feasible decisions. A policy for player q in an SG is exactly the same except that randomizations are permitted. Hence, a policy $\Pi_q = (\pi_{1q}, \pi_{2q}, \ldots, \pi_{Nq})$ for player q is a sequence (denumerable if $N = \infty$) such that $\pi_{nq}(H_n)$ is a probability distribution on A_s^q if s_n, the last element of H_n, specifies $s_n = s$.

DEFINITION 7. *A* policy $\Pi = (\Pi_q;\ q \varepsilon Q)$ *is a* Q-*tuple consisting of a randomized policy for player* q, *for each* $q \varepsilon Q$.

Our definition of policy is restricted to the *behavior strategies* of game theory. Briefly, a behavior strategy separates for each n the randomization for period n+1's decision from the randomizations for decisions in periods 1 through n. Example 5 below specifies a policy which is not a behavior strategy. The restriction to behavior strategies is without loss of optimality in the following sense. For the solution concepts in Section 2, if the other players are using behavior strategies, then you cannot do any better outside the class of behavior strategies than you can do within the class of behavior strategies.

Definition 7 admits policies for player q which are not Markov policies (cf. Definition IV.3.4). In words, a policy for player q permits the n-*th* decision to depend on more of the past history than merely the n-*th* state. Theorem 4 in Section 3 presents sufficient conditions for existence of an SG solution which *is* a Markov policy.

We amend the notation in (3a,b,c) to indicate the dependence on the players' policy: $V^q(\Pi,N)$, $V^q(\Pi)$, and $G^q(\Pi)$. Let lower case letters denote expected values[1] of these r.v.'s:

(4a) $v_s^q(\Pi,N) = E[V^q(\Pi,N)|s_1 = s]$,

(4b) $v_s^q(\Pi) = E[V^q(\Pi)|s_1 = s]$,

and

(4c) $g_s^q(\Pi) = E[G^q(\Pi)|s_1 = s]$.

For the remainder of the chapter, we revert to our customary use of **boldface type** to denote vectors.

EXAMPLE 1 (Tossing Nickels). Two gamblers, Mutt and Jeff, are going to play the following game. Mutt has two nickels. Nickel 1 is biased so that its probability of heads is 1/3; nickel 2 has probability 3/4 of heads. Mutt and Jeff each have five pennies. At each play of the game, Mutt chooses which of his nickels he will flip and, simultaneously, Jeff decides positive (+) or negative (-). Depending on whether the flipped nickel falls heads (H) or tails (T), and whether Jeff decided on + or -, one of the players gives a penny to

[1] As with MDP's, sufficiently general models admit policies Π for which $V^q(\Pi,N)$, $V^q(\Pi)$, or $G^q(\Pi)$ may not be bona fied r.v.'s or, if they are, for which the expectations may not exist.

the other according to Table 1.

		Outcome of Flip	
		Heads	Tails
Jeff's Choice	+	Mutt gives Jeff 1¢	Jeff gives Mutt 1¢
	-	Jeff gives Mutt 1¢	Mutt gives Jeff 1¢

Table 1. *Movements of the Nickel Tossing Game*

The game ends when one of the players has all ten pennies (nickels are not convertible to pennies).

Here, $\mathcal{Q} = \{1,2\}$; let q =1 label Mutt and q = 2 label Jeff. Let state s label Mutt's penny holdings so Jeff holds 10-s pennies and the initial state is $s_1 = 5$. We adjoin "dummy" states s = -1 and s = 11 to $\{0,1,\ldots, 10\}$ and let $S = \{-1,0,\ldots, 11\}$. State s = -1 indicates that the game ended with s = 0 so Mutt had no pennies and Jeff had ten. Similarly, s = 11 indicates termination with Mutt holding ten pennies. Mutt's actions are chosen from $A_s^1 = \{1,2\}$ where 1(2) indicates nickel 1(2), if $s\varepsilon\{1,2, \ldots, 9\}$. Similarly, Jeff's actions are chosen from $A_s^2 = \{+,-\}$ if $s\varepsilon\{1,\ldots, 9\}$. If $s\varepsilon\{-1,0,10,11\}$ let $A_s^1 = A_s^2 = \{0\}$ where "action 0" is a "dummy action."

Most of the transition probabilities are given below:

		Jeff's Action, A^2	
		+	-
Mutt's Action, A^1	1	2/3	1/3
	2	1/4	3/4

Table 2. $p_{s,s+1}(a^1,a^2) = 1 - p_{s,s-1}(a^1,a^2)$

Table 2 specifies $p_{s,s+1}(a) = 1 - p_{s,s-1}(a)$ if $s\varepsilon\{1, \ldots, 9\}$. For these values of s, $p_{sj}(a) = 0$ if $j\not\varepsilon\{s+1, s-1\}$. Finally, let $p_{0,-1}(0,0) = p_{-1,-1}(0,) = p_{10,11}(0,0) = p_{11,11}(0,0) = 1$ so states -1 and 11 are absorbing.

The single stage reward function is $r_1[10,(0,0)] = r_2[0,(0,0)] = 10$ and $r_q(s,a) = 0$ for all other values of q, s, and a. In words, Mutt keeps ten pennies if the game ends at s=10, and Jeff keeps ten pennies if the game ends at s=0. We use the artifice of $N = \infty$, i.e. an infinite duration, to overcome the difficulty due to the fact that the actual length of the game is an r.v. whose distribution depends on how the players play the game. $\square$

EXAMPLE 2 (Advertising and Pricing). Consider two retail stores which compete with one another primarily on the basis of price of goods sold and advertising. Suppose that wholesale costs are approximately the same for both stores and proportional to quantities sold. Let c denote the wholesale cost per unit sold. Suppose also that the stores experience negligible inventory costs because wholesalers are located close by.

Let $\rho_{n,q}$ and $z_{n,q}$ denote store q's price and advertising expenditure in period n, respectively. Let $\underset{\sim}{a}_{n,q} = (\rho_{n,q}, z_{n,q})$ and $\underset{\sim}{a}_n = (\underset{\sim}{a}_{n,1}, \underset{\sim}{a}_{n,2})$.

Suppose, as in Section 4 of Chapter III, that "goodwill" represents the impacts of advertising expenditures on demand. Let $s_{n,q}$ denote store q's goodwill at the beginning of period n; let $\underset{\sim}{s}_n = (s_{n,1}, s_{n,2})$. We assume that each store's goodwill in period n depends, perhaps probabilistically, on both stores' goodwills and advertising expenditures in period n-1. A simple specific model of goodwill that is similar to the model in Section 4 of Chapter III is

$$s_{n,q} = \theta_q(s_{n-1,q} + z_{n-1,q}), \qquad q = 1,\ 2,$$

where $0 < \theta_q < 1$. The parameter θ_q represents the rate at which store q's goodwill deteriorates.

Let $D_{n,q}$ denote the demand (in dollars) for store q's goods in period n; let $\underset{\sim}{D}_n = (D_{n,1}, D_{n,2})$. We assume that $\underset{\sim}{D}_n$ is a random vector with a distribution which depends only on s_n and $\underset{\sim}{a}_n$. One of several examples we shall analyze in §4 has $D_{n,q}$ uniformly distributed on $[0,\ (s_{n,q} + z_{n,q})/(s_{n,1} + s_{n,2} + z_{n,1} + z_{n,2})]$ if $s_{n,q} + z_{n,q} > 0$, and $P\{D_{n,q} = 0\} = 1$ if $s_{n,q} + z_{n,q} = 0$.

If each firm uses the discount factor β, then the sum of discounted gross variable profits for store q is

$$\Sigma_{n=1}^{\infty}\beta^{n-1}[(\rho_{n,q} - c)D_{n,q} - z_{n,q}]. \tag{5}$$

Let

$$\mu_q(s,a) = E(D_{n,q} | s_n = s,\ \underset{\sim}{a}_n = \underset{\sim}{a})$$

$$r_q(s,a) = (\rho_{n,q} - c)\mu_q(s,\underset{\sim}{a}) - z_q$$

where $\underset{\sim}{a} = ((\rho_1, z_1), (\rho_2, z_2))$. Then the expected value of (5) is the same as the expected value of

$$\Sigma_{n=1}^{\infty} \beta^{n-1} r_q(s_n, \underset{\sim}{a}_n).$$

This model satisfies the definition of an SG with $Q = \{1,2\}$, $S = \mathbb{R}_+^2$ (goodwill is scaled to be nonnegative), and $A_s^q = \mathbb{R}_+^2$ for each $s \varepsilon S$ and $q \varepsilon Q$. $\square$

EXAMPLE 3 (Competing Banks). Suppose that a town has three banks and that nearly all the residents maintain their checking accounts at these banks. Let $Q = \{1,2,3\}$ and let $s_{n,j}^j$ denote the fraction of the residents with accounts at bank j at the beginning of month n; let $\underset{\sim}{s}_n = (s_{n,1}, s_{n,2}, s_{n,3})$. Then $\underset{\sim}{s}_n \varepsilon S = \{(x_1, x_2, x_3)\colon x_i \geq 0 \text{ all } i \text{ and } \Sigma x_i = 1\}$.

Suppose that the banks do not engage in price competition but that they do vary in the lengths of time that customers must wait at tellers' windows. Let $a_{n,q}$ denote the number of tellers employed at bank j during month n; let $\underset{\sim}{a}_n = (a_{n,1}, a_{n,2}, a_{n,3})$. In the short-run, each bank is limited by its physical structure; let m_q denote the number of tellers' windows in bank q. Then $A_{\underset{\sim}{s}}^q = \{1, \ldots, m^q\}$ for each $\underset{\sim}{s}$ and q. A simple model for the dependence of the fraction of accounts on quality of service is

$$s_{n+1,q} = (1 - \gamma_q) s_{n,q} + \gamma_q a_{n,q} (a_{n,1} + a_{n,2} + a_{n,3})$$

where $0 < \gamma_q < 1$.

Suppose that the operating profit in a month depends on the amount of funds deposited in a bank and that the amount of funds is proportional to the fraction of residents with accounts at the bank. Then a reasonable model for the expected monthly operating profit is

$$r_q(\underset{\sim}{s},\underset{\sim}{a}) = w_q(s_q^q) - c_q(a_q)$$

where $w_q(\cdot)$ describes revenue due to invested assets and $c_q(\cdot)$ describes the cost of employing a_q tellers. $\square$

EXAMPLE 4 (Oligopoly with Pricing, Production, and Inventories). Suppose that a collection Q of manufacturers are competitors. Let $s_{n,q}$ denote the amount of the q-*th* firm's stock of finished goods at the beginning of period n; let $\underset{\sim}{s}_n = (s_{n,q};q\varepsilon Q)$ and $S = \mathbb{R}^Q$. Let $z_{n,q}$ denote the quantity of goods produced by the q-*th* firm during period n and let $y_{n,q} = s_{n,q} + z_{n,q}$. The constraint $z_{n,q} \geq 0$ implies $y_{n,q} \geq s_{n,q}$. Suppose that production is sufficiently rapid (relative to a period's length) so that $y_{n,q}$ is the total amount of goods available to satisfy demand during period n. Let $\underset{\sim}{y}_n = (y_{n,q};q\varepsilon Q)$.

Let $D_{n,q}$ denote firm q's demand during period n; let $\underset{\sim}{D}_n = (D_{n,q};q\varepsilon Q)$. Under the assumption that excess demand is backlogged, $s_{n+1,q} = y_{n,q} - D_{n,q}$ so

$$\underset{\sim}{s}_{n+1} = \underset{\sim}{y}_n - \underset{\sim}{D}_n. \tag{6}$$

Let $\rho_{n,q}$ denote the price charged by firm n during period n. Let $\underset{\sim}{a}_{n,q} = (y_{n,q},\rho_{n,q})$ and $\underset{\sim}{a}_n = (a_{n,q};q\varepsilon Q)$. Then $A_s^q = [s_q,\infty) \times [0,\infty)$. We assume that $\underset{\sim}{D}_1$ given $\underset{\sim}{a}_1$, $\underset{\sim}{D}_2$ given $\underset{\sim}{a}_2,\ldots$ is a sequence of conditionally independent and identically distributed random vectors. With (6), this assumption implies that the distribution of $\underset{\sim}{s}_{n+1}$ is entirely determined by $\underset{\sim}{a}_n$:

$$P\{\underset{\sim}{s}_{n+1} \leq \underset{\sim}{x} | \underset{\sim}{a}_n = \underset{\sim}{a}\} = P\{D_n \geq \underset{\sim}{a} - \underset{\sim}{x} | \underset{\sim}{a}_n = \underset{\sim}{a}\}$$

where $\underset{\sim}{x} \in \mathbb{R}^Q$.

We make revenue and cost assumptions similar to those in Section 2 of Chapter III for the monopoly case. Suppose that c_q is the q-*th* firm's unit cost of production so $z_{n,q}$ costs $c_q z_{n,q}$. Let $g_q(\underset{\sim}{y},\underset{\sim}{\rho},\underset{\sim}{d})$ denote the q-*th* firm's revenue minus inventory and stockout costs in any period in which $\underset{\sim}{y}_n = \underset{\sim}{y}$, $\underset{\sim}{\rho}_n = \underset{\sim}{\rho}$, and $\underset{\sim}{D}_n = \underset{\sim}{d}$. An example is

$$g_q(\underset{\sim}{y},\underset{\sim}{\rho},\underset{\sim}{d}) = -h_q \cdot (y_q - d_q)^+ - \pi_q \cdot (d_q - y_q)^+ + \rho_q d_q .$$

Recall that $\underset{\sim}{a}_{n,q} = (y_{n,q}, \rho_{n,q})$. Therefore, the specification of a model which satisfies the definition of an SG is completed with

$$r_q(\underset{\sim}{s},\underset{\sim}{a}) = E[g_q(\underset{\sim}{y},\underset{\sim}{\rho},\underset{\sim}{D}_1) | \underset{\sim}{a}_1 = \underset{\sim}{a}] - c_q \cdot (y_q - s_q)$$

where s_q and y_q are the q-*th* components of $\underset{\sim}{s}$ and $\underset{\sim}{y}$, respectively, and $\underset{\sim}{a}_1 = (a_{1,q};\ q \in Q)$. $\square$

EXAMPLE 5. Consider a two period MDP. This is an SG with $Q = \{1\}$ and $N = 2$. Let $S = \{1\}$ and $A_1^1 = \{1,2\}$. The set of all policies, according to Definition 7, is the set of all triples $(\alpha_1, \alpha_{21}, \alpha_{22})$ where α_1 is the probability of taking action 1 in period 1 and, for $j = 1$ or 2, α_{2j} is the probability of taking action 1 in period 2 if action j was taken in period 1. In each instance, one minus alpha is the probability of taking action 2.

Here is a decision rule which is not a *policy*, i.e. is not admitted by Definition 7. With probability 1/2, take action 1 in both periods; with probability 1/2 take action 2 in both periods. This rule requires a joint randomization which is precluded by Definition 7. $\square$

2. Solution Concepts

Specific sequential game models are descriptions of interactions among several persons or institutions. What is a "solution" of such a model? It can be construed as a prediction of the behavior that would actually occur in the modeled context. Also, it can be viewed as a recommendation of how the players ought to play the game. For either purpose, one must at least anticipate the "other" players' behavior. But behavior often seems to depend on the type of setting and personality as much as on the actual rewards and dynamics. This pluralism leads to numerous concepts of solution for a game. In this section, we define several concepts of solution for sequential games.

Suppose that all players know the initial state s_1. For much of this section we fix $s_1 = s$. Suppose further that each player $q \in Q$ chooses a criterion among (1.4a,b,c): either the finite horizon[1] expected discounted return $v_s^q(\Pi,N)$, the infinite horizon expected discounted return $v_s^q(\Pi)$, or the expected average return per period $g_s^q(\Pi)$. If each player q selects a policy Π_q, let $\Pi = (\Pi_q;\ q \in Q)$. Then Π induces a value for each player's criterion. Let $w_q(\Pi)$ denote the numerical value of player q's criterion if the players use policy Π (and s is the initial state).

EXAMPLE 1. Suppose that there are two players, that each has only two alternative policies, and that the resulting values of $w_q(\Pi)$ are given below. Let $c_1 \in \{1,2\}$ and $c_2 \in \{1,2\}$ label the policies of players 1 and 2, respectively. The entries in Table 1 are $[w_1(c_1,c_2),\ w_2(c_1,c_2)]$, i.e. the first entry in a cell is player 1's payoff and the second is player 2's.

---

[1]If *any* player $q \in Q$ uses the criterion $v_s^q(\Pi|N)$, then we assume that *all* players use that criterion with the duration N being the same integer from each player's perspective.

		Player 2's Policy c_2	
		1	2
Player 1's Policy, c_1	1	4,5	2,6
	2	5,3	3,2

Table 1. *Payoffs in a One Period Game:* $[w_1(c_1,c_2),\ w_2(c_1,c_2)]$

Suppose $(c_1,c_2) = (2,1)$. Then the payoffs are (5,3). If player 2 adheres to $c_2 = 1$, how attractive to player 1 is $c_1 = 2$ vs. $c_1 = 1$? Player 1's payoffs is 4 if $c_1 = 1$ versus 5 if $c_1 = 2$. We assume "more is better" so player 1 would prefer to use $c_1 = 2$ *if player 2 adheres to* $c_2 = 1$.

Similarly, if player 1 adheres to $c_1 = 2$, what should player 2 do? The comparison is a payoff of 3 via $c_2 = 1$ vs. 2 if $c_2 = 1$. Hence, player 2 would prefer to use $c_2 = 1$ *if player 2 adheres to* $c_1 = 2$. We say that $(c_1,c_2) = (2,1)$ is an *equilibrium point*. Neither player has an incentive to deviate from his or her portion of $(c_1,c_2) = (2,1)$ if the other player is steadfast. Each player checks, or balances, the other.

If the players use $(c_1,c_2) = (2,1)$, their payoffs sum to $5 + 3 = 8$. However, if they use $(c_1,c_2) = (1,1)$ then the sum is $4 + 5 = 9$ and no other policy leads to a higher sum. We say that $(c_1,c_2) = (1,1)$ is a *joint maximum*.

Suppose each player is trying to hurt the other one. Each strives to minimize the other's payoff. If player 1 uses $c_1 = 1$, then player 2, by using $c_2 = 2$, can limit player 1 to a payoff of 2. If player 1 uses $c_1 = 2$, then player 2, by using $c_2 = 2$, can limit player 1 to a payoff of 3. If player 2 uses $c_2 = 1$, player 1's best rejoinder (i.e. policy which minimizes player 2's payoff) is $c_1 = 2$. If player 2 uses $c_2 = 2$, player 1's best rejoinder is $c_1 = 2$. Therefore, each component of $(c_1,c_2) = (2,2)$ hurts the opponent as much as possible. This is an *ill-fare equilibrium point*. Note that $(c_1,c_2) = (2,2)$ is an equilibrium point for the following bimatrix game used in place of Table 1.

		c_2	
		1	2
c_1	1	-5,-4	-6,-2
	2	-3,-5	-2,-3

Table 2. *Payoffs for an Ill-Fare Equilibrium Point*

The entry at (c_1,c_2) in Table 2 is $[-w_2(c_1,c_2), -w_1(c_1,c_2)]$ where $w_1(c_1,c_2)$ and $w_2(c_1,c_2)$ are taken from Table 1.

Player 1 prefers $(c_1,c_2) = (2,1)$ to $(c_1,c_2) = (1,1)$ because the payoff would be 5 instead of 4. However, player 2's payoff would drop from 5 to 3. Similarly, player 2 prefers $(c_1,c_2) = (1,2)$ to (1,1) because the payoff would be 6 instead of 5. However, player 1's payoff would drop from 4 to 2. Therefore, $(c_1,c_2) = (1,1)$ has the property that neither player's welfare can be improved without injuring the other player. We say that $(c_1,c_2) = (1,1)$ is *Pareto optimal* (called *efficient* or *admissable* in some literatures). Also, (1,2) and (2,1) are Pareto optimal.

Suppose each player tries to maximize the amount by which his or her payoff is higher than the opponent's payoff. Then we replace the entries in Table 1 with $[w_1(c_1,c_2) - w_2(c_1,c_2), w_2(c_1,c_2) - w_1(c_1,c_2]$ as in Table 3.

		Player 2's Policy c_2	
		1	2
Player 1's Policy c_1	1	-1,1	-4,4
	2	2,-2	1,-1

Table 3. *Payoffs for the Criterion: Maxmin the Difference*

The entries in each cell then sum to zero. We call this criterion *maxmin the difference*. The policy $(c_1,c_2) = (2,2)$ is an equilibrium point in Table 3. That is, $w_1(\cdot,2) - w_2(\cdot,2)$ is maximized at $c_1 = 2$, and $w_2(2,\cdot) - w_1(2,\cdot)$ is maximized at $c_2=2$.

Suppose, instead, that each player tries to maximize the amount by which his or her payoff exceeds the average payoff. If $(c_1,c_2) = (1,1)$ then the payoffs are (4,5) so the average is 4.5 and, with respect to this average, the players' increments are (-.5,.5). If $(c_1,c_2) = (1,2)$, the increments are (-2,2); if $(c_1,c_2) = (2,1)$, the increments are (1,-1); and if $(c_1,c_2) = (2,2)$, the increments are (.5,-.5). Multiplying these increments by two yields the entries in Table 3. Therefore, (c_1,c_2) is an equilibrium point for the criterion maxmin the difference if, and only if, it is an equilibrium point for the criterion *beat the average*. If the game had more than two players, then beat the average would be well-defined whereas maxmin the difference would be ill-defined. $\square$

Equilibrium Point of a Static Game

In order to define an equilibrium point of an SG (sequential game) and prove that one exists, it is convenient first to define an equilibrium point of a single period SG. A single period SG is called a *static game, noncooperative game,* or *Nash game* (after John Nash who first proved existence of equilibrium points under reasonably general conditions).

DEFINITION 1. *A* static game *consists of a nonempty set* Q*, for each* $q \in Q$ *a nonempty set* W_q*, and for each* $q \in Q$ *a real-valued function* $m_q(\cdot)$ *defined on* $W = \times_{i \in Q} W_i$.

We interpret W_q as the set of *choices* available to player q. In some applications, W_q consists of the specific actions available to player q. In others, W_q is a set of probability distributions on the set of specific actions available to player q. The latter interpretation arises if we admit randomized strategies.

Suppose W_q is a set of randomized strategies and Y_q is the set of specific actions available to player q. If Y_q is a denumerable set, usually W_q is the set of *all* probability distributions on Y_q, i.e. the set of all $[y(q,k);\ k \varepsilon Y_q] \geq \underset{\sim}{0}$ such that

$$\Sigma_{k \varepsilon Y_q} y(q,k) = 1.$$

If there are more than denumerably many specific actions in Y_q, then (for technical reasons) W_q is a proper subset of the set of all probability distributions on Y_q. Where this issue might arise in SG's, we shall make an assumption comparable to $\#Y_q < \infty$ for each $q \varepsilon Q$.

Suppose for each $q \varepsilon Q$ that $w_q \varepsilon W_q$ is player q's choice, and let $\underset{\sim}{w} = (w_q;\ q \varepsilon Q)$. Sometimes it is convenient to separate a specific player's component in $\underset{\sim}{w}$ from those of the other players. Hence, we abuse notation and and for a specific $q \varepsilon Q$ write $\underset{\sim}{w} = (w_q, \underset{\sim}{w}_{-q})$ where $\underset{\sim}{w}_{-q}$ denotes the choices of all the players except player q. Similarly, for any $i \varepsilon Q$, $q \varepsilon Q$, and $\underset{\sim}{w} \varepsilon W$, we write $m_i(\underset{\sim}{w}) = m_i(w_q, \underset{\sim}{w}_{-q})$.

DEFINITION 2. *A static game has an* equilibrium point (EP) $\underset{\sim}{w}^*$ with respect to W *if*

$$(1) \qquad m_q(\underset{\sim}{w}^*) \geq m_q(w_q, \underset{\sim}{w}^*_{-q}) \quad \textit{for all} \quad w_q \varepsilon w_q \quad \textit{and} \quad q \varepsilon Q.$$

When W is clear from the context, we often write "EP" rather than "EP with respect to W." The qualification "with respect to W" is important. For example, in some static games, if W represents only unrandomized specific actions, then no EP exists. In many such games, if W is expanded to include randomizations of actions, then an EP exists. If a static game has an EP with respect to unrandomized specific actions then that multiplayer action remains an EP with respect to randomizations of specific actions.

From Appendix B, a set X is a convex and bounded polyhedron if X has only finitely many extreme points $x_1,\ldots,x_m$ and, for all $x \varepsilon X$, there are nonnegative numbers $\lambda_1,\ldots,\lambda_m$ such that

$$\Sigma_{i=1}^{m}\lambda_i = 1 \quad \text{and} \quad x = \Sigma_{i=1}^{m}\lambda_i\, x_i.$$

PROPOSITION 1 (John Nash). *If a static game has finitely many players and, for each* $q \varepsilon \mathcal{Q}$, $m_q(\cdot)$ *is continuous on* W *and* W_q *is a closed, convex, and bounded polyhedron, then there is an EP with respect to* W.

PROOF. John Nash (1950)[1]. $\square$

PROPOSITION 2 (John Nash). *Suppose a static game has finitely many players and, for each q,* W_q *consists of all the probability distributions on a finite set* Y_q *and* $m_q(\cdot)$ *is player q's expected payoff. Then there is an EP with respect to* W.

PROOF. We shall apply Proposition 1. For each q, W_q consists of the nonnegative solutions $[y(q,k);\ k\varepsilon Y_q]$ to $\Sigma_{k\varepsilon y_q} y(q,k) = 1$. This is a closed, bounded, and convex polyhedron.

Let $z_q(k_1,\ldots,k_Q)$ denote player q's payoff if the players take unrandomized actions $(k_1,\ldots,k_Q)$. If the players use $y = [y(q,k);\ k\varepsilon Y_q,\ q\varepsilon \mathcal{Q}]\varepsilon W$ then player q's expected payoff is

$$m_q(\underset{\sim}{y}) = \Sigma_{k_1\varepsilon Y_1} \cdots \Sigma_{k_Q\varepsilon Y_Q} \prod_{i\varepsilon \mathcal{Q}} y(i,k_i)\, z_q(k_1,\ldots,k_Q).$$

it follows that $m_q(\cdot)$ is continuous on W_i for every i, hence on W. Proposition 1 now yields existence of an EP. $\square$

---

[1]J. Nash, "Equilibrium Points in n-Person Games," Proceedings National Academy of Sciences U.S.A., Vol. 36, 48-49, 1950.

Definitions for Sequential Games

Now we define equilibrium points and Pareto optima of SG's (sequential games). The other solution concepts in Example 1 were equilibrium points of transformations of the original game.

It is convenient to write the policy $\Pi = (\Pi_q;\ q\varepsilon Q)$ as the pair (Π_k, Π_{-k}) for a specific player $k\varepsilon Q$. Then Π_{-k} represents the policies of all the players except player k. Recall the notation $v_s^q(\Pi,N)$, $v_s^q(\Pi)$, and $g_s^q(\Pi)$ (from (1.4a,b,c)) for the expected total reward, expected discounted reward, and expected average reward per period. Let S' be a subset of states, let $\underset{\sim}{\pi}_q$ be a subset of player q's policies, and let $\underset{\sim}{\pi} = \underset{q\varepsilon Q}{\times} \underset{\sim}{\pi}_q$.

DEFINITION 1. *A policy* $\Pi^* = (\Pi_q^*;\ q\varepsilon Q)$ *is an* N-period equilibrium point (EP) with respect to *initial states in* S' and *policies in* $\underset{\sim}{\pi}$ *if*

$$(2) \qquad v_s^q(\Pi^*,N) \geq v_s^q[(\xi_q, \Pi_{-q}^*),\ N] \quad \textit{for all } s\varepsilon S';\ \ \xi_q \varepsilon \underset{\sim}{\pi}_q,\ \textit{and } q\varepsilon Q.$$

A policy Π^* *is a* discounted equilibrium point (EP) with respect to S' and $\underset{\sim}{\pi}$ *if*

$$(3) \qquad v_s^q(\Pi^*) \geq v_s^q[(\xi_q, \Pi_{-q}^*)] \quad \textit{for all } s\varepsilon S',\ \ \xi_q \varepsilon \underset{\sim}{\pi}_q,\ \ \textit{and}\ \ q\varepsilon Q.$$

A policy Π^* *is an* average reward equilibrium point (EP) with respect to S' and $\underset{\sim}{\pi}$ *if*

$$(4) \qquad g_s^q(\Pi^*) \geq g_s^q[(\xi_q, \Pi_{-q}^*)] \quad \textit{for all } s\varepsilon S',\ \xi_q \varepsilon \underset{\sim}{\pi}_q,\ \ \textit{and}\ \ q\varepsilon Q.$$

Usually the type of equilibrium point, S', and $\underset{\sim}{\pi}$ are clear from the context so we say merely *equilibrium point* and use the abbreviation *EP*.

There are several reasons why sometimes S' and $\underset{\sim}{\pi}$ are proper subsets of S and the set of all policies, respectively. First, proper subsets may be necessary to ensure existence of the expectations $v_s^q(\Pi,N)$, $v_s^q(\Pi)$, and $g_s^q(\Pi)$ (and to ensure that they correspond to bona fide r.v.'s). Second, we may be able to prove existence of EP's only for proper subsets. Third, there may be an EP with special properties only for proper subsets. An example of the second reason occurs in Section 3 when existence of discounted EP's is proven at first with respect to $\underset{\sim}{\pi}$ as the set of stationary policies. An example of the third reason occurs in Section 4 where we study EP analogues of the myopic optima in Chapter III.

An EP of a sequential game is closely related to the idea of EP for static games. We use discounted EP's to illustrate the relationship. A collection of static games corresponds to an SG with $\beta_q < 1$ for each $q\varepsilon Q$. Suppose $\underset{\sim}{\pi} = {}_{q\varepsilon Q}^{\times}\pi_q$; let $W = \underset{\sim}{\pi}$ and $W_q = \underset{\sim}{\pi}_q$, $q\varepsilon Q$. Fix $s\varepsilon S$; for each $q\varepsilon Q$ and $\Pi\varepsilon\underset{\sim}{\pi}$, let $m_q(\Pi) = v_s^q(\Pi)$. Call this the *s-th static game*. There is an *s-th* static game for each $s\varepsilon S$.

THEOREM 1. *A policy Π^* is a discounted EP with respect to S' and $\underset{\sim}{\pi}$ if and only if Π^* is an* EP *with respect to* $W = \underset{\sim}{\pi}$ *for every s-th static game with* $s\varepsilon S'$.

PROOF. Exercise 2. $\square$

It does *not* follow from Theorem 1 that the issue of existence of a discounted EP can be settled by the straight-forward application of Nash's theorem (Proposition 1). Theorem 1 requires that the *same* policy Π^* be an EP with respect to W for *all* of the s-*th* static games with $s\varepsilon S'$. The optimum of a discounted infinite horizon MDP can be construed as the solution of a vector maximization problem. Similarly, a discounted EP can be regarded as the solution of a vector static EP problem. In both cases, the components of the vector correspond to possible initial states.

Exercise 2 asks you to specify static games whose EP corresponds to an N-period EP and other static games whose EP corresponds to an average reward

EP.

Some of the results in Section 7 concern Pareto optima of sequential games.

DEFINITION 2. Π^* *is* Pareto optimal with respect to $\underset{\sim}{\pi}$ *if* $\xi \varepsilon \underset{\sim}{\pi}$, $q \varepsilon \mathcal{Q}$, *and* $s \varepsilon S$ *with* $v_s^q(\Pi^*) < v_s^q(\xi)$ *implies existence of* $j \varepsilon \mathcal{Q}$ *and* $u \varepsilon S$ *with* $v_u^j(\Pi^*) > v_u^j(\xi)$.

The definition of Pareto optimality can be altered so that the criterion is expected total reward or expected average reward per period. Following the three parts of Definition 1, the formalities should be obvious.

We abbreviate Pareto optimality with *PO*. A policy Π^* is PO if no player can be made better off from some initial state without some player becoming worse off from some initial state. Either the players or the initial states (but not both) may be the same.

EXERCISES

1. Define each of the following solution concepts for a sequential game with the criterion of expected discounted reward.
   (a) Joint maximum.
   (b) Ill-fare equilibrium point.
   (c) Maxmin the difference (assuming $Q = 2$).
   (d) Beat the average equilibrium point.
2. (a) Prove Theorem 1.
   (b) State analogues of Theorem 1 for N-period EP's and average reward EP's.

3.* Existence of Equilibrium Points

An equilibrium point has the property that the players' strategies *balance* each other. No player's expected payoff can be raised if all the other players adhere to their components of the strategy. The players implicitly hold each other to the (joint) strategy. Appending the idea of balance to a mathematical model usually leads to fixed point theorems. Therefore, it is hardly surprising that numerous authors have used fixed point theorems[1] to prove the existence of EP (equilibrium point) solutions to sequential game models.

In this section, we present sufficient conditions for the existence of discounted, average reward, and N-period equilibrium points. For discounted EP's, the existence proof is a generalization of John Nash's use of the Brouwer Fixed Point Theorem in (static) game theory. An alternative proof could be given based on the Contraction Mapping Fixed Point Theorem (cf. Section 2 in Chapter V) and the Kakutani Fixed Point Theorem.

All existence proofs depend on the following observation. Suppose all the players but one are using stationary policies. Then the remaining player faces an MDP (Markov decision process) whose criterion is the expected present value of the rewards. Therefore, Corollary IV.2.1 asserts that the player can confine a search for an optimal Markov policy to the set of stationary policies. As a result, the material in Sections 4 and 5 of Chapter IV and Section 2 of Chapter V can be invoked in existence proofs.

This section considers only finite SG's, i.e. games with finitely many players, states, and actions. Here we introduce the notation which specifies the MDP

---

[1]The theorems invoked include the Contraction Mapping, Brouwer, and Kakutani Fixed Point Theorems and their corollaries. These may be found in references listed at the end of the chapter.

faced by one player when all the other players use stationary policies. From Section 2, recall that a policy is written $\Pi = (\Pi_q;\ q\varepsilon Q)$ with player q's policy being the component $\Pi_q = (\pi_{1q}, \pi_{2q}, \ldots)$ in which π_{nq} specifies player q's decision rule in the n-*th* period. The following definition of a single stage decision rule permits randomized rules. As Section 2 illustrates, even in static games there may not exist an EP unless randomized rules are permitted.

DEFINITION 1. *A* single stage decision rule for player q *is a collection* $\delta_q = \{\delta_q(s)\colon s\varepsilon S\}$ *such that* $\delta_q(s)$ *is a probability distribution on* A_s^q *for each* $s\varepsilon S$. *A* single stage decision rule δ *is a collection* $\{\delta_q\colon q\varepsilon Q\}$ *such that* δ_q *is a single stage decision rule for player* q, *for each* $q\varepsilon Q$. *If there is a single stage decision rule* δ_q *such that* $\pi_{nq}(H_n) = \delta_q(s_n)$ *for all* H_n *and* n, *then* Π_q *is a* stationary policy for player q. *The policy* $\Pi = (\Pi_q;\ q\varepsilon Q)$ *is a* stationary policy *if, for each* $q\varepsilon Q$, Π_q *is a stationary policy for player* q. *The symbols* $\underset{\sim}{Z}_q$ *and* $\underset{\sim}{Z}$ *denote the set of stationary policies for player q and the set of stationary policies, respectively.*

COMMENTS.

(i) From the definitions,

(1) $$\underset{\sim}{Z} = \times_{q \in \mathcal{Q}} \underset{\sim q}{Z}.$$

(ii) Recall that Z denotes the set of stationary policies in an MDP (Section 3 in Chapter IV).

(iii) Since the SG is finite, i.e. $\#\mathcal{G} < \infty$, there are only finitely many elements in the sets of players, states, and actions.

(iv) A single stage decision rule is a collection of probability distributions, one for each player in each state. The set A_s^q is finite and $k \varepsilon A_s^q$ so we interpret $\delta_q(s)(k)$ as the probability that player q will choose action k when the state is s.

A single stage decision rule corresponds to each stationary policy, and conversely. Therefore, we also use $\underset{\sim}{Z}$ and $\underset{\sim q}{Z}$ to denote the set of single stage decision rules and the set of single stage decision rules for player q, respectively. It will be clear from each context whether $\underset{\sim}{Z}$ ($\underset{\sim q}{Z}$) represents single stage decision rules (for player q) or stationary policies (for player q).

Let $m(s,q) = \#A_s^q$; $m(s,q) < \infty$ for each s and q in a finite SG. The set of all probability distributions on A_s^q is the set of all solutions of the system

(2) $$D_{sk}^q \geq 0, \quad k = 1, \ldots, m(s,q); \quad \Sigma_{k=1}^{m(s,q)} D_{sk}^q = 1.$$

Let $\underset{\sim sq}{Z}$ denote the set of all solutions to (2); then define

(3) $$\underset{\sim q}{Z} = \times_{s \varepsilon \mathcal{S}} \underset{\sim sq}{Z} \quad \text{and} \quad \underset{\sim}{Z} = \times_{q \varepsilon \mathcal{Q}} \underset{\sim q}{Z} = \times_{q \varepsilon \mathcal{Q}} \times_{s \varepsilon \mathcal{S}} \underset{\sim sq}{Z}.$$

Let $\underset{\sim}{\delta} = (\delta_q;\ q\varepsilon\mathcal{Q})\varepsilon\underset{\sim}{Z}$. Then for each $s\varepsilon S$ and $q\varepsilon\mathcal{Q}$, $\delta_q(s)$ is specified by exactly one solution to (2). Let $\{D^q_{sk}:\ k\varepsilon A^q_s,\ q\varepsilon\mathcal{Q},\ s\varepsilon S\}$ correspond to $\underset{\sim}{\delta}$. Recall that we use Q to denote the number of players, i.e. the size of $\mathcal{Q}$.

Let

$$(4)\qquad b_{sj}(\underset{\sim}{\delta}) = \Sigma^{m(s,1)}_{k_1=1} \cdots \Sigma^{m(s,Q)}_{k_Q=1}\ p_{sj}(k_1,\ldots,k_Q)\ \Pi_{q\varepsilon\mathcal{Q}}\ D^q_{sk_q}$$

which is the expected value of the transition probability induced by $\underset{\sim}{\delta}$. Let $B_{\underset{\sim}{\delta}}$ represent the matrix of these transition probabilities and let $\underset{\sim}{\delta}^\infty$ indicate the stationary policy in which player q uses the policy δ^∞_q, for each $q\varepsilon\mathcal{Q}$.

Let $\underset{\sim}{\rho}^q(\underset{\sim}{\delta})$ be the vector whose s-*th* component is

$$(5)\qquad \rho^q_s(\underset{\sim}{\delta}) = \Sigma^{m(s,1)}_{k_1=1} \cdots\cdots \Sigma^{m(s,Q)}_{k_Q=1}\ \Pi_{i\varepsilon\mathcal{Q}}\ D^i_{sk}\ r_q(s,k_1,\ldots,k_Q).$$

This quantity is player q's expected single stage reward induced by policy $\underset{\sim}{\delta}^\infty$ whenever the state is s. Let $v^q_s(\underset{\sim}{\delta})$ be player q's expected present value induced by policy $\underset{\sim}{\delta}^\infty$ from the initial state s; let $\underset{\sim}{v}^q(\underset{\sim}{\delta})$ denote the vector whose s-*th* component is $v^q_s(\underset{\sim}{\delta})$. Then $\beta_q < 1$ and Theorem IV.5.1 (alternatively, Theorem V.2.1) imply

$$(6)\qquad \underset{\sim}{v}^q(\underset{\sim}{\delta}) = \Sigma^\infty_{i=0}(\beta_q B_{\underset{\sim}{\delta}})^i\ \underset{\sim}{\rho}^q(\underset{\sim}{\delta}) = (I - \beta_q B_{\underset{\sim}{\delta}})^{-1}\ \underset{\sim}{\rho}^q(\underset{\sim}{\delta})$$

(where $B^0_{\underset{\sim}{\delta}} = I$).

Existence of Discounted EP's

The main result, Theorem 1 below, asserts that a finite SG necessarily has a discounted EP among the class of stationary policies. The proof has two parts.

First, we construct a mapping τ and use Brouwer's fixed point theorem to prove that τ has a nonempty set of fixed points. Second, we show that the fixed points of τ are necessarily EP's and conversely.

DEFINITION 2. *A function* f *with domain and range* $\mathcal{D}$ *has* fixed point d *if* $d\varepsilon\mathcal{D}$ *and* $f(d) = d$.

We abbreviate *fixed point* with *FP*. Brouwer's theorem presents sufficient conditions for the existence of an FP in $\mathcal{D}$.

PROPOSITION 1 (Brouwer's Fixed Point Theorem). *Let* f *be a function with domain and range* $\mathcal{D}$. *If* $\mathcal{D}$ *is a closed, bounded, and convex subset of* $\mathbb{R}^m$ $(m < \infty)$, *and if* f *is continuous on* $\mathcal{D}$, *then* f *has an FP in* $\mathcal{D}$.

THEOREM 1. *If a finite SG has* $\beta_q < 1$ *for all* $q\varepsilon\mathcal{Q}$ *then there exists a discounted EP with respect to* S *and* $\underset{\sim}{Z}$.

PROOF. Step (i) constructs a mapping τ on $\underset{\sim}{Z}$ and uses Brouwer's theorem to establish that τ has an FP in $\underset{\sim}{Z}$. Step (ii) proves that $\underset{\sim}{\delta}\varepsilon\underset{\sim}{Z}$ is an FP of τ if, and only if, $\underset{\sim}{\delta}^\infty$ is an EP.

(i) In order to use Brouwer's theorem, $\underset{\sim}{Z}$ must be a closed, bounded, and convex subset of $\mathbb{R}^m$ for some $m < \infty$. If $\underset{\sim}{Z}_q$ has the required properties for each $q\varepsilon\mathcal{Q}$, then $\underset{\sim}{Z} = \times_{q\varepsilon\mathcal{Q}} \underset{\sim}{Z}_q$ has them too. The SG is finite so $m(s,q) = \# A_s^q < \infty$ for each s and q. Using (2) and (3), it is simple to verify that $\underset{\sim}{Z}_{sq}$ is a closed, bounded, and convex subset of $\mathbb{R}^{m(s,q)}$ (Exercise 1 asks you to do this). Let

$$m \triangleq \Sigma_{s\varepsilon S}\, \Sigma_{q\varepsilon\mathcal{Q}}\, m(s,q).$$

Then $\underset{\sim}{Z} \subset \mathbb{R}^m$, $m < \infty$, and $\underset{\sim}{Z}$ is closed, bounded, and convex.

Let $\underset{\sim}{\delta}_{sk}^q$ denote the modification of $\underset{\sim}{\delta}$ in which player q takes action $k\varepsilon A_s^q$ with probability one if the state is s; i.e. $D_{sk}^q = 1$ and $D_{sj}^q = 0$ if $j \neq k$. Then

$$\rho_s^q(\underset{\sim}{\delta}_{sk}^q) + \beta_q \, \Sigma_{j\varepsilon S} b_{sj}(\underset{\sim}{\delta}_{sk}^q) v_j^q(\underset{\sim}{\delta})$$

is player q's expected discounted return from the initial state s if all the other players adhere to their portions of $\underset{\sim}{\delta}^\infty$ while player q uses $\underset{\sim}{\delta}_{sk}^q$ in the first period and, thereafter, δ_q^∞. Let

$$(7) \qquad \phi_{sk}^q(\underset{\sim}{\delta}) = [\rho_s^q(\underset{\sim}{\delta}_{sk}^q) + \beta_q \, \Sigma_{j\varepsilon S} b_{sj}(\underset{\sim}{\delta}_{sk}^q) v_j^q(\underset{\sim}{\delta}) - v_s^q(\underset{\sim}{\delta})]^+$$

indicate the increase in player q's expected discounted return, if any, from deferring δ_q^∞ for one period during which $\underset{\sim}{\delta}_{sk}^q$ is used. Define a function $\tau\colon \underset{\sim}{Z} \to \underset{\sim}{Z}$ with

$$(8) \qquad \tau(\underset{\sim}{\delta})(s,k,q) = \frac{D_{sk}^q + \phi_{sk}^q(\underset{\sim}{\delta})}{1 + \Sigma_{i=1}^{m(s,q)} \; \phi_{si}^q(\underset{\sim}{\delta})}, \quad k\varepsilon A_s^q, \quad s\varepsilon S, \; q\varepsilon \mathcal{Q}.$$

Exercise 2 asks you to verify that $\tau(\underset{\sim}{\delta})\varepsilon \underset{\sim}{Z}$ and that $\tau(\cdot)$ is continuous on $\underset{\sim}{Z}$. Therefore, Brouwer's theorem implies existence of an FP of τ.

(ii) $\{EP\} \subset \{FP\}$: Suppose that $\underset{\sim}{\delta}^\infty$ is an EP and all players except q use their components of $\underset{\sim}{\delta}^\infty$. Then player q faces an MDP whose transition probabilities and rewards are specified by

$$\{b_{sj}(\underset{\sim}{\delta}_{sk}^q)\colon k\varepsilon A_s^q,\ s\varepsilon S\} \quad \text{and} \quad \{\rho_s^q(\underset{\sim}{\delta}_{sk}^q)\colon k\varepsilon A_s^q,\ s\varepsilon S\},$$

respectively. The definition of an EP (Definition 2.1) implies that δ_q^∞ is an optimal policy for this MDP. Then Theorem IV.5.3 implies $\phi_{sk}^q(\underset{\sim}{\delta}) = 0$ for all s, k, and q. Therefore, $\tau(\underset{\sim}{\delta}) = \underset{\sim}{\delta}$ from (8) so $\underset{\sim}{\delta}$ is an FP.

{EP} $\supset$ {FP}: Suppose δ is an FP so $\tau(\underset{\sim}{\delta}) = \underset{\sim}{\delta}$. We shall prove $\phi_{sk}^{q}(\underset{\sim}{\delta}) = 0$ for all s, k, and q so, by Theorem IV.5.3, $\underset{\sim}{\delta}^{\infty}$ is an EP. From (8),

$$D_{sk}^{q} = \frac{D_{sk}^{q} + \phi_{sk}^{q}(\underset{\sim}{\delta})}{1 + \Sigma_{i=1}^{m(s,q)} \phi_{si}^{q}(\underset{\sim}{\delta})}, \quad k \varepsilon A_s^q, \quad s \varepsilon S, \quad q \varepsilon Q,$$

so

$$\text{(9)} \qquad \phi_{sk}^{q} = D_{sk}^{q} \, \Sigma_{i=1}^{m(s,q)} \, \phi_{si}^{q}, \quad k \varepsilon A_s^q, \quad s \varepsilon S, \quad q \varepsilon Q,$$

where we suppress the notational dependence of ϕ_{si}^{q} on $\underset{\sim}{\delta}$. From (9), if $D_{sk}^{q} = 0$ then $\phi_{sk}^{q} = 0$. Suppose, for some q and s,

$$\Sigma_{j=1}^{m(s,q)} \, \phi_{sj}^{q} > 0.$$

Then

$$D_{sk}^{q} = \phi_{sk}^{q} / \Sigma_{j=1}^{m(s,q)} \phi_{sj}^{q}, \qquad k \varepsilon A_s^q,$$

so $D_{sk}^{q} > 0$ implies $\phi_{sk}^{q} > 0$. Also $D_{sk}^{q} > 0$ for some $k \varepsilon A_s^q$ from (2). But we shall prove that $\phi_{sk}^{q} = 0$ for some k such that $D_{sk}^{q} > 0$. From (4), (5), and (6),

$$v_s^q(\underset{\sim}{\delta}) = \Sigma_{k=1}^{m(s,q)} D_{sk}^{q} v_{sk}^{q}(\underset{\sim}{\delta})$$

$$\geq \min\{v_{sk}^{q}(\underset{\sim}{\delta}) : k \varepsilon A_s^q \quad \text{and} \quad D_{sk}^{q} > 0\}.$$

Let j be a value of k at which min {...} is attained so $v_{sj}^q(\delta) - v_s^q(\delta) \leq 0$ while $D_{sj}^q > 0$. Therefore, $\phi_{sj}^q(\delta) = 0$ while $D_{sj}^q > 0$. Hence, $\phi_{sj}^q(\delta) = 0$ for all s, j, and q so δ^∞ is an EP. $\square$

Under the assumptions of Theorem 1, if δ^∞ is an EP then no player has an incentive to move to another *stationary* policy if the other players adhere to their portions of δ^∞. The following result shows that no player can obtain an improvement by choosing a Markov policy (Definition IV.3.4) that is not stationary, i.e. outside $\underset{\sim}{Z}$.

COROLLARY 1.1. *If an SG is finite and* $\beta_q < 1$ *for each* $q \varepsilon \mathcal{Q}$, *then there is a stationary policy which is a discounted EP with respect to S and the set of all Markov policies.*

PROOF. From Theorem 1, there exists a stationary policy δ^∞ which is an EP with respect to S and $\underset{\sim}{Z}$. For any player q, if all the players but q use δ_{-q}^∞, their portion of δ^∞, then player q faces an MDP which satisfies the conditions of Corollary IV.2.1. That result asserts that an MDP has an optimal stationary policy if it has an optimal Markov policy. Therefore, within the class of Markov policies player q cannot improve upon δ_q^∞ as a rejoinder to δ_{-q}^∞. This is true of all players $q \varepsilon \mathcal{Q}$. $\square$

Existence of an Average Reward Equilibrium Point

From definition (2.1), $\Pi^* = (\Pi_q^*;\ q \varepsilon \mathcal{Q})$ is an average reward equilibrium point with respect to initial states in S' and policies in $\underset{\sim}{\pi}$ if

$$(10) \qquad g_s^q(\Pi^*) \geq g_s^q[(\xi_q, \Pi_{-q}^*)] \text{ for all } s \varepsilon S', \quad \xi \varepsilon \pi, \quad \text{and} \quad q \varepsilon \mathcal{Q}.$$

The search for Π^* to satisfy (10) is at least as hard as the search for an optimal policy in an MDP with the criterion of average return. From Section 6 in Chapter

IV, the MDP problem is simpler if every stationary policy induces a Markov chain with only one communicating class of states. We make the corresponding assumption below. As you might guess from Section 3 of Chapter V, the existence of an equilibrium point can be proved under weaker conditions.[1]

UNICHAIN ASSUMPTION. *Every single stage decision rule* δ *has a transition matrix* B_δ *that induces a Markov chain with one communicating class of states and a (possible empty) set of transient states.*

The unichain assumption and finitely many states implies (Section 6 of Chapter VII in Volume 1) existence of a stationary distribution which is the same regardless of the initial state. That is, for each δ there is a probability vector $\underset{\sim}{c}^{\delta}$ with components c_s^{δ}, $s \in S$, such that

$$(11) \qquad \underset{\sim}{c}(\delta) \geq \underset{\sim}{0}, \quad \underset{\sim}{c}(\delta) \cdot \underset{\sim}{e} = 1, \quad \text{and} \quad \underset{\sim}{c}(\delta) = \underset{\sim}{c}(\delta) B(\delta),$$

where $\underset{\sim}{e}$ denotes the column vector whose components are all one. The importance of the unichain assumption in the following proof is that it simplifies the verification that τ is continuous on $\underset{\sim}{Z}$. *Some* assumption is needed (Exercise 3).

THEOREM 2. *Suppose a finite SG satisfies the unichain assumption. Then there exists an average reward EP with respect to S and* $\underset{\sim}{Z}$.

PROOF. The proof is similar to that of Theorem 1. We already know that $\underset{\sim}{Z}$ is a closed, bounded, and convex subset of $\mathbb{R}^m$ with $m < \infty$.

For each δ, the unichain assumption and Theorem IV.6.1 (cf (17) through (22) in Section 6 of Chapter IV) imply for each $q \in Q$ that there exists a unique number $g^q(\delta)$

---

[1]A. Federgruen, "Successive Approximation Methods in Undiscounted Stochastic Games," Operations Research, Vol. 28, 794-809, 1980.

and vector $\underset{\sim}{w}^q(\delta)$ with components $w_s^q(\delta)$ such that

$$(12) \quad g^q(\delta) = \underset{\sim}{c}(\delta)\cdot\underset{\sim}{\rho}_\delta^q$$

$$(13) \quad \underset{\sim}{c}(\delta)\cdot\underset{\sim}{w}^q(\delta) = 0$$

$$(14) \quad \underset{\sim}{e}\cdot g^q(\delta) + \underset{\sim}{w}^q(\delta) = \underset{\sim}{\rho}_\delta^q + B_\delta \underset{\sim}{w}^q(\delta)$$

where $\underset{\sim}{c}(\delta)$ is the stationary distribution which uniquely solves (11). Let

$$(15) \quad \phi_{sk}^q(\delta) = [\rho_s^q(\delta_{sk}^q) + \Sigma_{j\varepsilon S} b_{sj}(\delta_{sk}^q) w_j^q(\delta) - g^q(\delta) - w_s^q(\delta)]^+$$

which is similar to the test quantity (IV.6-43) in the policy improvement algorithm for the average return criterion under the unichain assumption. Let $\{D_{sk}^q\}$ correspond to δ. Then define τ: $\underset{\sim}{Z} \to \underset{\sim}{Z}$ with

$$\tau(\delta)(s,k,q) = \frac{D_{sk}^q + \phi_{sk}^q(\delta)}{1 + \Sigma_{i=1}^{m(s,q)} \phi_{si}^q(\delta)}, \quad k\varepsilon A_s^q, \quad s\varepsilon S, \quad q\varepsilon Q,$$

which is the same as (8) except that $\phi_{sk}^q(\delta)$ is specified by (15) instead of (7).

Continuity of τ on $\underset{\sim}{Z}$ will be shown below; then Brouwer's theorem implies existence of an FP. To show that an FP δ is necessarily an EP, we have as in the proof of Theorem 1 that $\phi_{sk}^q(\delta) = 0$ for all s, k, and q. Therefore, due to Theorem IV.6.3, δ^∞ is an EP. Conversely, if δ^∞ is an EP then all $\phi_{sk}^q(\delta) = 0$ due to Theorem IV.6.3 so δ is an FP of τ.

Continuity of τ on $\underset{\sim}{Z}$ will follow from continuity of $g^q(\cdot)$ and $\underset{\sim}{w}^q(\cdot)$ for each $q\varepsilon Q$. If $g^q(\cdot)$, B_δ, and $\underset{\sim}{c}(\cdot)$ are continuous then so also is $\underset{\sim}{w}^q(\cdot)$ due to (13) and

(14). For continuity of $g^q(\cdot)$, it is sufficient to prove that B_δ and $\underset{\sim}{c}(\cdot)$ are continuous and to verify that B_δ has one communicating class of states (and, perhaps, some transient states).

Recall that $\underset{\sim}{Z}$ is a convex set and let H denote its set of extreme points; $\#H < \infty$. Hence, $\delta\varepsilon\underset{\sim}{Z}$, implies existence of a finite set $\{\alpha_i : i\varepsilon H\}$ such that $\alpha_i \geq 0$ for all i, $\Sigma_{i\varepsilon H}\, \alpha_i = 1$, and

$$(16) \quad B_\delta = \Sigma_{i\varepsilon H}\, \alpha_i\, P_i$$

where P_i is the matrix of transition probabilities under the i-*th* extreme point policy. It follows from (16) that $B_{\underset{\sim}{\delta}}$ is continuous on $\underset{\sim}{Z}$. As a consequence of Exercises 4 and 5, $B_{\underset{\sim}{\delta}}$ has one communicating class and it and $\underset{\sim}{c}(\cdot)$ are continuous on $\underset{\sim}{Z}$. $\square$

Existence of an N-Period Equilibrium Point

From Section 1, recall the notation

$$(17a) \quad V^q(\underset{\sim}{\Pi}, N) = \Sigma_{n=1}^{N} \beta_q^{n-1}\, r_q(s_n, \underset{\sim}{a}_n) + \beta_q^N L_q(s_{N+1})$$

and

$$(17b) \quad v_s^q(\underset{\sim}{\Pi}, N) = E[V^q(\underset{\sim}{\Pi}, N) \mid s_1 = s]$$

which makes explicit the dependence on the players' policy $\underset{\sim}{\Pi} = (\Pi_q;\ q\varepsilon\mathcal{Q})$. From Definition 2.1, a policy $\underset{\sim}{\Pi}^*$ is an N-period EP with respect to initial states in S' and policies in $\underset{\sim}{\pi} = \times_{q\varepsilon\mathcal{Q}}\, \underset{\sim}{\pi}_q$ if

$$(18) \qquad v_s^q(\underset{\sim}{\Pi}^*, N) \geq v_s^q[(\xi_q, \underset{\sim}{\Pi}^*_{-q}), N], \quad s\varepsilon S', \quad \xi_q \varepsilon \pi_q, \quad \text{and} \quad q\varepsilon\mathcal{Q}.$$

Recall the notation in Definitions 2.1 and 2.2 of a static game and an EP of a static game. From Exercise 2.2(b), Π^* is an N-period EP with respect to S' and $\underset{\sim}{\pi}$ if and only if it is an EP with respect to $W = \underset{\sim}{\pi}$ for every static game, called the *s-th static game,* obtained by fixing $s \in S'$ and letting $W_q = \underset{\sim}{\pi}_q$ and $m_q(\Pi) = v_s^q(\Pi,N)$, $q \in Q$.

THEOREM 3. *For every* $N \in I_+$, *a finite SG has an N-period EP with respect to S and the set of all policies.*

PROOF. Fix $N \in I_+$. We shall construct a policy which is simultaneously an EP for every s-*th* static game, $s \in S$. Fix $s \in S$. A policy for player q is a sequence $\Pi_q = (\pi_{1q}, \pi_{2q}, \ldots, \pi_{Nq})$ with $\pi_{nq}(H_n)$ being a probability distribution on $A^q_{s_n}$ (where s_n is the last element of history H_n). For fixed n, a finite SG can have only finitely many sequences $H_n = (s_1, a_1, \ldots, s_{n-1}, a_{n-1}, s_n)$, hence only finitely many sequences $(H_1, \ldots, H_N)$, hence only finitely many unrandomized sequences $\Pi_q = (\pi_{1q}, \ldots, \pi_{Nq})$. Label player q's unrandomized policies $\Pi^1_q, \ldots, \Pi^{d(q)}_q$ with $d(q) < \infty$. Exercise 8 asks you to prove that any Π_q is a convex combination of these unrandomized policies. Therefore, the set of all policies for player q is a closed, bounded, and convex polyhedron. For each q, player q's payoff in the s-*th* static game is the expected payoff. Therefore, Proposition 2.2 implies existence of an EP $\Pi^*(s)$ in the s-*th* static game.

Continue to keep $s \in S$ fixed. Since $\Pi^*(s)$ is a bona fide policy, it stipulates a randomized action a_n for each possible history H_n, for n=1,...,N. The possible histories include all possible initial states but the payoffs $v^q_s(\Pi,N)$ in the s-*th* static game depend only on randomized actions in response to histories whose initial state is s. Let $\Pi^*(s)_j$ denote the portion of $\Pi^*(s)_j$ which stipulates randomized actions $a_1, \ldots, a_N$ for histories with initial state $j \in S$. Therefore, in $\Pi^*(s)$, if $j \neq s$, we may alter $\Pi^*(s)_j$ and the altered $\Pi^*(s)$ will remain an EP in the s-*th* static game.

Now let s vary in S and construct a policy with components $\Pi^*(s)_s$, $s\varepsilon S$. For each $s\varepsilon S$, if the initial state is s, then the randomized actions are part of an EP in the s-*th* static game. From the argument above, such a constructed policy is an EP in every s-*th* static game, $s\varepsilon S$, hence an N-period EP. $\square$

Recall the discussion below Definition 1.7 which observes that our definition of "policy" admits non-Markov decision rules.

DEFINITION 3. *A* Markov policy $\Pi = (\Pi_q; q\varepsilon Q)$ *in an N-period SG is a policy in which player q's policy* $\Pi_q = (\pi_{1q},\ldots,\pi_{Nq})$ *has* π_{nq}, *for each n, depend on* H_n *only through its last element,* s_n.

It follows from the definition that a Markov policy is a sequence $(\delta^1,\delta^2,\ldots,\delta^N)$ in which δ^n depends only on the state when n periods remain until the game ends.

The construction in Theorem 3's proof does not yield any insight into the structure of an N-period EP. The following proof is essentially an alternative proof of Theorem 3 which shows that there is an N-period EP which is a Markov policy. Thus there is an EP Π^* in which player q's component Π_q^* consists of a sequence $(\delta_q^1,\ldots,\delta_q^N)$ where, for each n, $\delta_q^n(s)$ is a probability distribution on A_s^q. We interpret $\delta_q^n(s)_k$ as the probability that player q will take action k in period n if $s_n = s$.

THEOREM 4. *For every* $N\varepsilon I_+$, *a finite SG has a Markov policy which is an N-period EP with respect to S and the set of all policies.*

PROOF. Let V be the set of all real-valued functions on $S\times Q$ and let

$$(19) \qquad h(s,a,q,v) = r_q(s,a) + \beta_q \textstyle\sum_{j\varepsilon S} p_{sj}(a)v(j,q), \quad (s,a)\varepsilon \mathcal{J},\ q\varepsilon Q, \text{ and } v\varepsilon V.$$

To initiate an inductive proof, let N=1. Then Theorem 3 asserts[1] existence of a 1-period EP δ^1 with respect to S and the set of all policies. Trivially, δ^1 is a Markov policy.

Suppose the theorem is valid for N-1. From the comment below Definition 3, there is an N-1-period EP $(\delta^{N-1},\delta^{N-2},\ldots,\delta^2,\delta^1)$ in which δ^n is the decision rule which determines the randomized action when n periods remain until the game ends. Also, δ^n depends only on the state at the beginning of the period, i.e. s_{N-n+1}.

Let $v(j,q) = v_j^q[(\delta^{N-1},\ldots,\delta^1),N]$ which is player q's expected return in the N-1-period game, from initial state j, if the players use the N-1-period EP $(\delta^{N-1},\ldots,\delta^1)$. Then h(s,a,q,v) in (19) is player q's expected return in the N-period game in which the initial state is s, the players take action a, and then ıse the N-1-period EP $(\delta^{N-1},\ldots,\delta^1)$. Fix $s\varepsilon S$ and let $m_q(a) = h(s,a,q,v)$. The SG s finite so there are only finitely many $a\varepsilon \times_{q\varepsilon\Omega} A_s^q$. Consider the static game in hich each player q's payoff is the expected value of $m_q(\cdot)$ and in which player q ıy randomize on A_s^q. From Proposition 2.2, this static game has an EP $\delta^N(s)$. Let $:= [\delta^N(s); s\varepsilon S]$. We shall prove that $(\delta^N,\delta^{N-1},\ldots,\delta^1)$ is an N-period EP.

We write $V_s^q(\Pi,N)$ for the r.v. $V^q(\Pi,N)$ in (17a) if s_1=s and E_Π r an expectation with respect to probabilities induced by policy Π. Let z^{N-1} ıote $(\delta^{N-1},\ldots,\delta^1)$ and $z^N = (\delta^N,z^{N-1})$; let $v(j,q) = v_j^q(z^{N-1}, N-1)$. Suppose all players but q use policy $z_{-q}^N = (\delta_{-q}^N,z_{-q}^{N-1})$ and player q uses an N-period policy Then player q's expected return from initial state s is

---

[1]Alternatively, for N=1 appeal directly to Proposition 2.2 in the proof of rem 3 with N=1.

$$v_s^q[(\xi_q,z_{-q}^N),N] = E_{(\xi_q,z_{-q}^N)}\{r_q(s,a) + \beta_q\Sigma_{j\varepsilon S}\ p_{sj}(a)v_j^q[(\xi_q,z_{-q}^N),\ N-1]\}$$

$$= E_{(\xi_q,z_{-q}^N)}\{r_q(s,a) + \beta_q\Sigma_{j\varepsilon S}\ p_{sj}(a)v_j^q[(\xi_q,z_{-q}^{N-1}),N-1]\}$$

$$\leq E_{(\xi_q,z_{-q}^N)}[r_q(s,a) + \beta_q\ \Sigma_{j\varepsilon S}\ p_{sj}(a)v_j^q(z^{N-1},N-1)] \tag{20}$$

$$= E_{(\xi_q,z_{-q}^N)}[h(s,a,q,v)] \leq E_{\delta N}[h(s,a,q,v)] \tag{21}$$

where a is the random action with N periods remaining. The inequality in (20) is due to z^{N-1} being an N-1 period EP. The inequality in (21) is caused by the EP property of δ^N and $z^N = (\delta^N,z^{N-1})$. Therefore, $z^N = (\delta^N,\ldots,\delta^1)$ is an N-period EP. $\square$

EXERCISES

1. (a) Using notation in part (i) of the proof of Theorem 1, prove that $\underset{\sim}{Z}_{sq}$ is a closed, bounded, and convex subset of $\mathbb{R}^{m(s,q)}$.

   (b) Use the result in (a) to prove that $\underset{\sim}{Z}$ is a closed, bounded, and convex subset of $\mathbb{R}^m$ where

$$m \triangleq \Sigma_{s\varepsilon S}\Sigma_{q\varepsilon Q}\, m(s,q).$$

2. (a) Using notation in part (i) of the proof of Theorem 1, prove $\underset{\sim}{\delta}\varepsilon\underset{\sim}{Z}$ implies $\tau(\underset{\sim}{\delta})\varepsilon\underset{\sim}{Z}$.

   (b) Prove that $\tau(\cdot)$ is continuous on $\underset{\sim}{Z}$. (Hint: use (4) through (8) to prove that $\phi^q_{sk}(\cdot)$, hence τ, is continuous).

3. Construct an example, with at most three states and two players, where τ constructed in the proof of Theorem 2 is *not* continuous on $\underset{\sim}{Z}$. (The example, of course, will not satisfy the hypothesis of Theorem 2).

4. Let $P_1,P_2,\ldots,P_m$ be a finite number of stochastic matrices of the same size. Suppose that P_1 induces a Markov chain with one communicating class (and, perhaps, some transient states). Let $h_1,h_2,\ldots,h_m$ be nonnegative numbers which sum to one and $h_1 > 0$. Prove the following properties of

(17) $$M = \Sigma^m_{i=1}h_iP_i.$$

   (a) If states k and j communicate in P_1 then they communicate in M.

   (b) M induces a Markov chain with one communicating class of states (and, perhaps, some transient states). (Hint for both parts: $M^n = \Sigma^m_{i=1}h^n_iP^n_i + \ldots. = h^n_1P^n_1 + \ldots.$)

5. In Exercise 4, suppose $P_1, P_2, \ldots, P_m$ each induce a Markov chain with one communicating class (and, perhaps, some transient states); the classes may differ from one matrix to the next. Prove the following properties of M in (17) without requiring $h_1 > 0$.

   (a) M induces a Markov chain with one communicating class of states (and, perhaps, some transient states).

   (b) Let H be the set of all $\underset{\sim}{h} = (h_1, \ldots, h_m) \geq \underset{\sim}{0}$ such that $\Sigma_{i=1}^{m} h_i = 1$. Let $M(\underset{\sim}{h})$ make the dependence on $\underset{\sim}{h}$ explicit. From (a), for each $\underset{\sim}{h} \varepsilon H$, there exists a unique solution $\underset{\sim}{c}(\underset{\sim}{h})$ to $\underset{\sim}{c}(\underset{\sim}{h}) \geq \underset{\sim}{0}$, $\underset{\sim}{c}(\underset{\sim}{h}) \cdot \underset{\sim}{e} = 1$, and $\underset{\sim}{c}(h) = \underset{\sim}{c}(\underset{\sim}{h}) M(\underset{\sim}{h})$. Then $\underset{\sim}{c}(\cdot)$ is continuous on H. (Hint: Suppose N(x) is a nonsingular matrix with finitely many rows for all $x \varepsilon X$ and $N(\cdot)$ is continuous on X; then $[N(\cdot)]^{-1}$ is continuous on X.)

. This exercise is along the lines of Section 2 of Chapter V. Suppose $\mathcal{S}$ and $\mathcal{Q}$ are countable and A_s^q is countable for each s and q. Let V be the set of all bounded real-valued functions on $\mathcal{S} \times \mathcal{Q}$ and let

$$d(u,v) = \sup\{|u(s,q) - v(s,q)| : s \varepsilon \mathcal{S},\ q \varepsilon \mathcal{Q}\}, \quad u \varepsilon V,\ v \varepsilon V.$$

   (a) Prove that V with $d(\cdot,\cdot)$ is a complete metric space.

   (b) Let $h(s,a,q,v) = r_q(s,a) + \beta_q \, \Sigma_{j \varepsilon \mathcal{S}} \, p_{sj}^{a} v(j,q)$ for $(s,a) \varepsilon \mathcal{G}$, $q \varepsilon \mathcal{Q}$, and $v \varepsilon V$. Assume that there exists $\beta < 1$ and $u < \infty$ such that $\beta_q \leq \beta$ for all $q \varepsilon \mathcal{Q}$ and $|r_q(s,a)| \leq u$ for all $q \varepsilon \mathcal{Q}$ and $(s,a) \varepsilon \mathcal{G}$. State analogues of the Contraction and Boundedness Assumptions in Section 2 of Chapter V, and prove that $h(\cdot,\cdot,\cdot,\cdot)$ satisfies these assumptions.

   (c) For each single stage decision rule δ and $v \varepsilon V$, let $H_\delta v$ be the mapping on $\mathcal{S} \times \mathcal{Q}$ which assigns to (s,q) the value $h[s,\delta(s),q,v]$. Prove for each δ that H_δ has a unique fixed point $v_\delta \varepsilon V$, and

$$d(v_{\underset{\sim}{\delta}},v) \le d(H_{\underset{\sim}{\delta}}v,v)/(1-\beta), \qquad v\varepsilon V.$$

(d) For each single stage decision rule $\underset{\sim}{\delta}$ and $v\varepsilon V$, let $L_{\underset{\sim}{\delta}}v$ be the following mapping on $S\times Q$:

$$[L_{\underset{\sim}{\delta}}v](s,q) = \sup\{h[s,a_q,\underset{\sim}{\delta}_{-q}(s),q,v]: a_q\varepsilon A_s^q\}, \qquad (s,q)\varepsilon S\times Q.$$

Prove that $L_{\underset{\sim}{\delta}}$ is a contraction mapping on V.

(e) Prove that h satisfies an analogue of the Monotonicity Assumption in Section 2 of Chapter V. Let $F_{\underset{\sim}{\delta}}$ be the unique FP in V of $L_{\underset{\sim}{\delta}}$, and define $f_{\underset{\sim}{\delta}}$ on $S\times Q$ as

$$f_{\underset{\sim}{\delta}}(s,q) = \sup\{v_{(\gamma,\underset{\sim}{\delta}_{-q})}(s,q): \gamma\varepsilon\Delta_q\}$$

where Δ_q is player q's set of single stage decision rules, and $v_{\underset{\sim}{\delta}}$ is the unique FP of $H_{\underset{\sim}{\delta}}$. Prove that $f_{\underset{\sim}{\delta}} = F_{\underset{\sim}{\delta}}$ for each single stage decision rule $\underset{\sim}{\delta}$.

7. The steps in Exercise 6 do *not* prove existence of an EP. For that result, $v_{\underset{\sim}{\delta}} = f_{\underset{\sim}{\delta}}$ for some $\underset{\sim}{\delta}$ is needed; then $\underset{\sim}{\delta}^\infty$ is an EP. This last step has been proven[1] using the Kakutani fixed point theorem. However, the problem is simpler if, as we assume now, there are only two players, rewards are zero-sum, i.e. $r_1(s,\underset{\sim}{a}) + r_2(s,\underset{\sim}{a}) = 0$ for all $(s,\underset{\sim}{a})\varepsilon \mathcal{G}$, and $\#\mathcal{G} < \infty$. Define h as in Exercise 6; here $h(s,\underset{\sim}{a},1,v) = -h(s,\underset{\sim}{a},2,v)$ for all $(s,\underset{\sim}{a})\varepsilon \mathcal{G}$ and $v\varepsilon V$ so, in the remainder of the exercise, let $h(\cdot,\cdot,\cdot) = h(\cdot,\cdot,1,\cdot)$. From the minimax theorem for matrix games, for each $v\varepsilon V$, define a mapping Gv on S via

[1] A. M. Fink, "Equilibrium in a Stochastic n-Person Game," Journal of Science of Hiroshima University, Series A-I, Vol. 26, 89-93, 1964.

$$[Gv](s) = \sup\{\inf\{h(s,\delta_1,\delta_2,v):\ \delta_2\varepsilon\Delta_2\}:\ \ \delta_1\varepsilon\Delta_1\}$$

$$= \inf\{\sup\{h(s,\delta_1,\delta_2,v):\ \delta_1\varepsilon\Delta_1\}:\ \ \delta_2\varepsilon\Delta_2\}, \qquad s\varepsilon S.$$

For each $\underset{\sim}{\delta} = (\delta_1,\delta_2)$ and $v\varepsilon V$, define a mapping $H_{\underset{\sim}{\delta}}v$ on S with $[H_{\underset{\sim}{\delta}}v](s) = h[s,\underset{\sim}{\delta}(s),v]$.

(a) Explain why an FP of G is an EP.

(b) Prove that $H_{\underset{\sim}{\delta}}$ is a contraction mapping, hence it has a unique FP $v_{\underset{\sim}{\delta}}$.

(c) Prove that G is a contraction mapping, hence it has a unique FP. (Hint: For each δ_1 and v let

$$[\mathscr{H}_{\delta_1}v](s) = \inf\{[H_{(\delta_1,\delta_2)}v](s):\ \delta_2\varepsilon\Delta_2\},$$

and prove that $\mathscr{H}_{\delta_1}$ is a contraction mapping. Note that

$$[Gv](s) = \sup\{[\mathscr{H}_{\delta_1}v](s):\ \delta_1\varepsilon\Delta_1\}$$

and use Theorem V.2.2.

(d) Let $v^0\varepsilon V$ be arbitrary and for each $i\varepsilon I_+$ let

$$v^i = Gv^{i-1} = G^i v^0$$

where $G^1 = G$ and $G^{i+1} = GG^i$. From (c), $\lim_{i\to\infty} v^i$ exists and is the FP of G. Fix i and explain (in detail), how to compute v^i starting from v^0.

8. Fix $N\varepsilon I_+$. In the proof of Theorem 3, it is claimed that any Π_q is a convex combination of the unrandomized Π_q's. Prove it.

9. Prove that $m_q(\cdot)$, defined in the proof of Theorem 4, is continuous on W (for each N, s, and q).

10. Theorems 3 and 4 are alternative existence proofs for N-period EP's. Construct a third proof by invoking Theorem 1. (Hint: For N fixed, construct an infinite horizon game with state space $S\times\{1,\ldots,N+1\}$.)

4. Myopic Equilibrium Points

Section 3 establishes that every finite SG, for every initial state, has an N-period EP (equilibrium point) for every N. Also, it has a discounted EP and, under certain conditions, it has an average reward EP. Section 8 contains algorithms for the computation and approximation of EP's but those algorithms are lengthy. This section presents sufficient conditions for the computation of an EP in an SG to be replaced by the computation of an EP in a static game. The latter task is much easier than the former.

The results in this section are analogous to Chapter III's simplification for MDP's. We find in Chapter III that myopic optima facilitate the qualitative analysis of optimal policies. In an example at the end of this section and throughout Section 5, we find that myopic EP's simplify the analysis of qualitative properties of EP's of SG's.

From section 3 of Chapter III, an MDP is said to have a *myopic optimum* if its data can be used easily to specify a single period optimization problem with the following property: *ad infinitum* repetition of a solution to the single period problem comprises an optimal MDP solution. Similarly, an SG is said to have a *myopic EP* if its data can be used easily to specify a static game with the following property: *ad infinitum* repetition of an EP of the static game comprises an EP for the SG.

The following assumptions are similar to Assumptions I through IV in Section 3 of Chapter III. It is convenient to define

$$A^q = \bigcup_{s \in S} A^q_s, \quad q \in Q, \qquad A = \times_{q \in Q} A^q$$

and

$$S(\underset{\sim}{a}) = \{s\colon (s,\underset{\sim}{a}) \varepsilon \mathcal{J}\}, \qquad \underset{\sim}{a} \varepsilon A.$$

Hence, $S(\underset{\sim}{a})$ is the set of states from which the multiplayer action $\underset{\sim}{a}$ is feasible. Recall Definition 1.4 which introduces the notation $L_q(s)$ for player q's salvage value if the ultimate state is s. Finally, in notation similar to that in Section 2, we write $\underset{\sim}{d} = (d_q, \underset{\sim}{d}_{-q}) \varepsilon A$ where d_q is player q's action and $\underset{\sim}{d}_{-q}$ denotes the actions of all the players except player q.

ASSUMPTION I:

$$(1) \qquad r_q(s,\underset{\sim}{a}) = K_q(\underset{\sim}{a}) + L_q(s), \qquad (s,\underset{\sim}{a}) \varepsilon \mathcal{J}, \quad q \varepsilon \mathcal{Q}.$$

ASSUMPTION II:

The transition function satisfies

$$(2) \qquad p(J|s,\underset{\sim}{a}) = x(J|\underset{\sim}{a}), \quad (s,\underset{\sim}{a}) \varepsilon \mathcal{J}, \quad \text{so } s_{n+1} \sim \xi(a_n), \quad n \varepsilon I_+.$$

Let

$$(3) \qquad \gamma_q(\underset{\sim}{a}) = K_q(\underset{\sim}{a}) + \beta_q E\{L_q[\xi(\underset{\sim}{a})]\}, \qquad \underset{\sim}{a} \varepsilon A,$$

and let Γ denote the following static game among the players in $\mathcal{Q}$. Player q has available the set of moves A^q and $\gamma_q(a)$ is player q's payoff when the players choose $a \varepsilon A$. First, suppose Γ has an EP in pure (i.e. unrandomized) strategies.

ASSUMPTION III:

There exists $\underset{\sim}{a}^* \varepsilon A$ *such that*

$$(4) \qquad \gamma_q(\underset{\sim}{a}^*) \geq \gamma_q(k, \underset{\sim}{a}^*_{-q}), \qquad k \varepsilon A^q, \qquad q \varepsilon \mathcal{Q}.$$

ASSUMPTION IV:[1]

$$\text{(5)} \qquad P\{\xi(a^*)\varepsilon S(a^*)\} = 1.$$

The MDP version of these assumptions is discussed in Section 3 of Chapter III. In particular, the transition function $p(\cdot|s,a)$, which ordinarily depends on both s and a, is assumed in (2) to depend only on a. As a result, the n+1-*st* state s_{n+1}, which ordinarily depends on both the n-*th* state s_n and the n-*th* multiplayer action a_n, is assumed in (2) to depend only on a_n. Therefore, s_{n+1} has the same probability distribution as a random variable $\xi(a_n)$ whose distribution, from (2), is

$$x(J|a) = P\{\xi(a)\varepsilon J\}.$$

Following Corollary 1.1 below, we relax Assumption III which requires Γ to have an EP in pure strategies.

Let α be a single stage decision rule which specifies $\alpha(s) = a^*$ (with probability one) if $s\varepsilon S(a^*)$ ($\Leftrightarrow a^*\varepsilon \times_{q\varepsilon\Omega} A_s^q$) and specifies an arbitrary element of $\times_{q\varepsilon\Omega} A_s^q$ if $s\notin S(a^*)$. Let α^N denote the N-period policy where $a_n = \alpha(s_n)$ for $n = 1, \ldots, N$.

THEOREM 1. *Assumptions* I *through* IV *imply*:

(a) α^∞ *is a discounted EP with respect to* $S(a^*)$ *and the set of all policies;*

(b) α^N *is an N-period EP, for every* $N\varepsilon I_+$, *with respect to* $S(a^*)$ *and the set of all policies.*

---

[1]Let $(\Omega,\mathcal{B},P)$ be the probability space here. By (5), we mean $P\{\xi(a^*)\varepsilon H\} = 1$ for all $H\varepsilon\mathcal{B}$ with $S(a^*)\subset H$.

PROOF. As in the proof of Theorem III.3.1, the substitution of (1) and (2) in

$$V^q(N) = \Sigma_{n=1}^{N}\beta_q^{n-1}r_q(s_n,a_n) + \beta_q^N L_q(s_{N+1}) \tag{6}$$

yields

$$V^q(N) = L^q(s_1) + \Sigma_{n=1}^{N}\beta_q^{n-1}\{K_q(a_n) + \beta_q L_q[\xi(a_n)]\}$$

so

$$E[V^q(N)] = L_q(s_1) + E[\Sigma_{n=1}^{N}\beta_q^{n-1}\gamma_q(a_n)] \tag{7}$$

where $\gamma_q(\cdot)$ is defined by (3).

Suppose $s_1 = s \varepsilon S(a^*)$ and all the players except player q use the single stage decision rule α_{-q}^N in periods 1,...,N. Let E_Π denote an expectation evaluated with probabilities induced by policy Π. If player q uses the N-period policy ξ, then q's expected return is

$$v_s^q[(\xi,\alpha_{-q}^N),N] = L_q(s) + E_{(\xi,\alpha_{-q}^N)}[\Sigma_{n=1}^{N}\beta_q^{n-1}\gamma_q(a_n)]$$

$$= L_q(s) + E_\xi[\,{}_{n=1}^{N}\beta_q^{n-1}\gamma_q(a_n^q,a_{-q}^*)]$$

$$\leq L_q(s) + \gamma_q(a^*)\Sigma_{n=1}^{N}\beta_q^{n-1} \tag{8}$$

with the inequality due to (4), the static EP property of a^*. Player q attains the right side of (8) by using the q-*th* component of a^* in every period. This conclusion is valid for all $q\varepsilon\mathcal{Q}$ which completes the proof of (b).

For (a), let $N \to \infty$ above. The EP claim is valid with respect to *all* policies, rather than only the Markov policies, for the following reason. From (4), if all the players but q are using α^∞_{-q} and player q uses *any* policy ξ, then

$$v^q_s(\xi,\alpha^\infty_{-q}) = L_q(s) + E_{(\xi,\alpha^\infty_{-q})}[\Sigma^\infty_{n=1}\beta_q^{n-1}\gamma_q(a^q_n,a^*_{-q})]$$

$$\leq L_q(s) + \gamma_q(a^*)/(1 - \beta). \tag{9}$$

Player q attains the right side of (9) by using the q-*th* component of α^∞. $\square$

COROLLARY 1.1 *Suppose there is* $u < \infty$ *such that*

$$|L_q(s)| \leq u, \qquad s\varepsilon S, \qquad q\varepsilon Q, \tag{10}$$

and Assumptions I *through* IV *are satisfied (with* $\beta_q = 1$ *in* (3)). *Then* α^∞ *is an average reward EP with respect to* $S(a^*)$ *and all policies.*

PROOF. Let (6) denote $V^q(N)$ with $\beta_q = 1$. Then

$$\Sigma^N_{n=1}r(s_n,a_n) = V^q(N) - L_q(s_{N+1})$$

so

$$G^q = \lim\inf_{N\to\infty}\Sigma^N_{n=1}r(s_n,a_n)/N$$

$$= \lim\inf_{N\to\infty}V^q(N)/N - \lim\inf_{N\to\infty}L_q(s_{N+1})/N. \tag{11}$$

From Theorem 1, if $s_1 \varepsilon S(\underset{\sim}{a}^*)$ then for each fixed N, each player's component of $\underset{\sim}{a}_n = \underset{\sim}{\alpha}(s_n)$, n=1,...,N, is an optimal rejoinder if all the other players do likewise. I.e.,

$$(12) \qquad \gamma_q(\underset{\sim}{a}^*) = v_s^q(\underset{\sim}{\alpha}^\infty, N)/N \geq v_s^q[(\xi_q, \underset{\sim}{\alpha}_{-q}^\infty), N]/N, \qquad s\varepsilon S(\underset{\sim}{a}^*),$$

for all policies ξ_q available to player q. Also, (10) implies $L_q(s_{N+1})/N \to 0$ for all policies. From (11) and (12), taking expected values in (12) and letting $N \to \infty$,

$$g_s^q(\underset{\sim}{\alpha}^\infty) = \gamma_q(\underset{\sim}{a}^*) \geq g_s^q[(\xi_q, \underset{\sim}{\alpha}_{-q}^\infty)], \qquad s\varepsilon S(\underset{\sim}{a}^*),$$

for all ξ_q. Hence, $\underset{\sim}{\alpha}^\infty$ satisfies (2.3), the definition of an average reward EP. $\square$

Recall from Section 3 where $\# A_s^q < \infty$ that $\underset{\sim}{Z}_{sq}$ denotes the set of all probability distributions on A_s^q. If A_s^q is not denumerable, then technical difficulties require a modification of the definition of $\underset{\sim}{Z}_{sq}$. For the remainder of this section, we avoid such modifications by considering only finite SG's. Now suppose Γ lacks a pure EP so Assumption III is not valid. Since the SG is finite, it follows that $A = \times_{q\varepsilon \underset{\sim}{Q}} \bigcup_{s\varepsilon S} A_s^q$ is a finite set so there necessarily exists at least a randomized EP (due to Proposition 2.2).

Formally,

$$\underset{\sim}{Z}^q = \bigcup_{s\varepsilon S} \underset{\sim}{Z}_{sq}, \qquad q\varepsilon \underset{\sim}{Q}, \qquad \text{and} \qquad \times_{q\varepsilon \underset{\sim}{Q}} \underset{\sim}{Z}^q$$

correspond to the sets A^q and A defined above Assumption I. In fact, $\underset{\sim}{Z}^q$ is contained in set of all probability distributions on A^q:

$$\underset{\sim}{Z}^q \subset \{(z_{iq}; i \varepsilon A^q): \quad z_{iq} \geq 0 \text{ for all } i \varepsilon A^q$$

$$\text{and } \Sigma_{i \varepsilon A^q} z_{iq} = 1\}.$$

We write $\underset{\sim}{z}_q \varepsilon \underset{\sim}{Z}^q$ for an element of $\underset{\sim}{Z}^q$. Then $\times_{q \varepsilon \mathcal{Q}} \underset{\sim}{Z}^q$ is the set of all $\underset{\sim}{z} = (\underset{\sim}{z}_q;\ q \varepsilon \mathcal{Q})$.
If the players in Γ use the randomized strategy $\underset{\sim}{z}$, then player q's expected return is

$$\rho^q(\underset{\sim}{z}) = \Sigma_{i_1 \varepsilon A_1} \cdots \Sigma_{i_Q \varepsilon A^Q} \prod_{j \varepsilon \mathcal{Q}} z_{i_j j} r_q(i_1, \ldots, i_Q)$$

which extends the notation (3.5). An EP $\underset{\sim}{z}$ of Γ, which necessarily exists, satisfies

$$\text{(13)} \qquad \rho^q(\underset{\sim}{z}) \geq \rho^q(\underset{\sim}{\xi}, \underset{\sim}{z}_{-q}), \qquad \underset{\sim}{\xi} \varepsilon \underset{\sim}{Z}^q, \quad q \varepsilon \mathcal{Q}.$$

For a particular EP $\underset{\sim}{z}$ of Γ, let A' denote the elements of A which are given positive probability:

$$A' = \{(k_1, \ldots, k_Q): (k_1, \ldots, k_Q) \varepsilon A \text{ and } z_{k_q q} > 0 \text{ for all } q\}.$$

Let S' denote the set of states from which every element of A' is feasible:

$$S' = \{s: s \varepsilon \mathcal{S} \quad \text{and} \quad a \ \varepsilon \times_{q \varepsilon \mathcal{Q}} A^q_s \qquad \text{for all } a \varepsilon A'\}$$

$$= \bigcap_{a \varepsilon A'} S(a).$$

The SG is assumed to be finite, so let $p_{sj}(a)$ denote the transition probability $r(\{j\}|s,a)$. Assumption II is equivalent to the existence of numbers $x_j(a)$, $j\varepsilon S$, $a\varepsilon A$, such that

$$x_j(a) = p_{sj}(a), \qquad (s,a)\varepsilon \mathcal{S},\ j\varepsilon S.$$

In place of Assumption IV, there is

ASSUMPTION IV':

(14) $\qquad \Sigma_{j\varepsilon S'}\, x_j(a) = 1 \qquad$ *for all* $a\varepsilon A'$.

The elements of A' are exactly the actions which are given positive probability by $\underset{\sim}{z}$. Then (14) asserts that, with probability one, the subsequent state will be in S'. But S' is exactly the set of states from which all the elements of A' are feasible.

Let λ be a single stage decision rule which uses $\underset{\sim}{z}$ if $s\varepsilon S'$ and is arbitrary if $s\notin S'$. Let λ^N denote the N-period policy which uses the single stage decision rule λ in periods 1,..., N.

THEOREM 2. *Suppose a finite SG satisfies assumptions I, II, and IV'.*

(a) λ^∞ *is a discounted EP and an average reward EP with respect to* S' *and all policies.*

(b) *For all* $N\varepsilon I_+$, λ^N *is an N-period EP with respect to* S' *and all policies.*

PROOF. Exercise 1. $\square$

Computation

Suppose an SG satisfies Assumptions I through IV or IV'. It follows from Theorems 1 and 2 that an EP solution of the SG is equivalent to finding an EP of the static game Γ specified between (3) and (4). Suppose that the SG is finite so that each player in Γ is randomizing over only finitely many actions, i.e. A^q is finite for each q. The Bibliographic Guide at the end of this chapter includes references to several algorithms which may be used to compute or approximate an EP of Γ.

A Useful Special Case

Instances of the following structure arise later in this section and in Sections 5 and 6. It is convenient to write $\underset{\sim}{a} = (a_1, a_2, \ldots, a_Q)$ instead of $(a^1, a^2, \ldots, a^Q)$ in stating and proving the following result.

THEOREM 3. *Suppose* $Q \geq 2$ *and for all* $q \in \mathcal{Q}$ *that* $\gamma_q(\underset{\sim}{a})$ *in* (3) *takes the form*

$$\gamma_q(\underset{\sim}{a}) = \alpha_q + \omega_q a_q / \Sigma_{i\in\mathcal{Q}}\, a_i - \xi_q a_q$$

with $\omega_q > 0$ *and* $\xi_q \geq 0$. *Let* $\eta_q = \xi_q/\omega_q$ *and suppose* $\Sigma_{i\in\mathcal{Q}}\, \eta_i \neq 0$. *Let* $\sigma = (Q-1)/\Sigma_{i\in\mathcal{Q}}\, \eta_i$ *and* $\underset{\sim}{a}^* = (a_q^*;\ q\in\mathcal{Q})$ *where*

$$a_q^* = \sigma - \eta_q\, \sigma^2, \qquad q\in\mathcal{Q}.$$

If $\eta_q \leq \sigma^{-1}$ *for each* q, $\Sigma_{i\in\mathcal{Q}}\ a_i^* \neq 0$, *and* $\underset{\sim}{a}^* \in A$ *then* $\underset{\sim}{a}^*$ *is an EP of* Γ *(the static game specified below* (3)). *If also* $\eta_q = \eta$ *for all* $q\in\mathcal{Q}$ *then*

$$a_q^* = (Q-1)/(Q^2\eta), \qquad q\in\mathcal{Q}.$$

PROOF. The assumed structure of $\gamma_q(\cdot)$ yields

$$\partial\gamma_q(\underset{\sim}{a})/\partial a_q = \omega_q \Sigma_{i\neq q}\ a_i/(\Sigma_{i\in\mathcal{Q}} a_i)^2 - \xi_q$$

$$\partial^2\gamma_q(\underset{\sim}{a})/\partial a_q^2 = -2\omega_q(\Sigma_{i\neq q} a_i)/(\Sigma_{i\in\mathcal{Q}} a_i)^3.$$

The assumptions imply $a_i^* \geq 0$ for each i so the second derivative is nonpositive at $\underset{\sim}{a} = \underset{\sim}{a}^*$. Setting the first derivative equal to zero yields

(15) $$\Sigma_{i\varepsilon \underset{\sim}{Q}}\, a_i = a_q + \eta_q(\Sigma_{i\varepsilon \underset{\sim}{Q}}\, a_i)^2.$$

Sum both sides over $q\varepsilon \underset{\sim}{Q}$ to obtain

$$Q\, \Sigma_{i\varepsilon \underset{\sim}{Q}}\, a_i = \Sigma_{q\varepsilon \underset{\sim}{Q}}\, a_q + (\Sigma_{q\varepsilon \underset{\sim}{Q}}\eta_q)(\Sigma_{i\varepsilon \underset{\sim}{Q}}\, a_i)^2.$$

Dividing both sides by $\Sigma_{i\varepsilon \underset{\sim}{Q}}\, a_i$ ($\neq 0$) leads to

$$\Sigma_{i\varepsilon \underset{\sim}{Q}}\, a_i = (Q-1)/\Sigma_{i\varepsilon \underset{\sim}{Q}}\eta_i = \sigma$$

which, upon substitution in (15), yields $a_q = \sigma - \eta_q\sigma^2$. If $\eta_q = \eta$ for all q then $\sigma = (Q-1)/(Q\eta)$ so

$$a_q = \sigma - \eta_q\sigma^2 = [(Q-1)/(Q\eta)][1 - (Q-1)/Q]$$

$$= (Q-1)/(Q^2\eta). \qquad \square$$

"Guns or Butter" Example

Interactions among nations and among biological species are sometimes discussed in terms of tradeoffs between consumption and investment. The idea is that a nation (species) is more likely to survive in the future if it invests more and consumes less. However, one of the purposes of survival is to be able to consume, so one would rather not forego consumption. These considerations are implicit in the following model. It can be regarded as a sequential game generalization of the resource management and capital accumulation model in Section 4 of Chapter VII.

Suppose there are two players whose wealths at the beginning of period n are (s_n^1, s_n^2). Each period n, player q decides how much to invest in "guns", g_n^q, and how much to consume as "butter," b_n^q. A player cannot use more than the available wealth so $g_n^q + b_n^q \le s_n^q$ and

$$A^q_{(s^1,s^2)} = \{(g,b)\colon\ 0 \le g,\ 0 \le b, \text{ and } g + b \le s^q\};$$

for each q, $0 \le s^q$.

Let (X_1^1, X_1^2), (X_2^1, X_2^2), (X_3^1, X_3^2),...be independent and identically distributed random vectors. We assume

$$\text{(16)} \qquad s_{n+1}^q = \begin{cases} X_n^q[s_n^q - b_n^q - g_n^q + h_q(\dfrac{g_n^q}{g_n^1 + g_n^2})] & \text{if } g_n^q > 0 \\ \\ X_n^q(s_n^q - b_n^q) & \text{if } g_n^q = 0 \end{cases}$$

where $h_q > 0$. In (16), $s_n^q - b_n^q - g_n^q$ is the residual wealth after consumption and investment, and $h_q g_n^q/(g_n^1 + g_n^2)$ is proportional to player q's fraction of the total investment. Hence, s^q_{n+1} is a random multiple of the sum of these two amounts. The term $h_q g^q_n/(g^1_n + g^2_n)$ has the property that the effectiveness of player q's investment in "guns" diminishes as the other player's investment in "guns" increases.

Suppose that player q's reward in period n is the amount of "butter" that q consumes, b^q_n. Then

$$(17) \qquad r_q\{(s^1,s^2),\ [(g^1,b^1),\ (g^2,b^2)]\} = b^q, \qquad q = 1,2.$$

This model satisfies Assumption I because, in (17), $r_q(s,\underset{\sim}{a}) = K_q(\underset{\sim}{a}) + L_q(s)$ with $L_q(\cdot) \equiv 0$ and $K_q(\underset{\sim}{a}) = K_q[(g^1,b^1),\ (g^2,b^2)] = b^q$. However, Assumption II is violated by (16).

Assumption II stipulates that s_{n+1} depends only on a_n. Therefore, we transform the definition of player q's action from (g^q,b^q) to $(s^q - b^q,g^q)$ and replace A^q_s in (13) with its equivalent for the new action (d^q,g^q) where $d^q = s^q - b^q$:

$$A^q_{(s^1,s^2)} = \{(d,g):\ \ 0 \leq g \leq d \leq s^q\}.$$

Recall the notation

$$A^q = \bigcup_{s \in S} A^q_s \ .$$

Here, $S = \mathbb{R}^2_+$ so

$$(18) \qquad A^q = \{(d,g):\ \ 0 \leq g \leq d\}, \qquad q = 1,2.$$

Also, $S(\underset{\sim}{a})$ denotes $\{s: (s,a) \varepsilon \mathcal{S}\}$ so (17) yields

$$(19) \qquad S[(d^1,g^1), (d^2,g^2)] = \{(s^1,s^2): \; 0 \leq g^q \leq d^q \leq s^q, \quad q = 1,2\}.$$

Let $\underset{\sim}{a}^q$ now denote the new action (d^q,g^q), let $\underset{\sim}{a} = (a^1,a^2)$, and let $\underset{\sim}{s} = (s^1,s^2)$. Then (19) becomes

$$r_q(\underset{\sim}{s},\underset{\sim}{a}) = s^q - d^q$$

because $s^q - d^q = s^q - (s^q - b^q) = b^q$. In this form, $r_q(\cdot,\cdot)$ satisfies Assumption I with $K_q(\underset{\sim}{a}) = -d^q$ and $L_q(\underset{\sim}{s}) = s^q$. Also, (16) becomes

$$(20) \qquad s^q_{n+1} = \begin{cases} X^q_n\left[d^q_n - g^q_n + h_q\left(\dfrac{g^q_n}{g^1_n + g^2_n}\right)\right] & \text{if } g^q_n > 0 \\ X^q_n d^q_n & \text{if } g^q_n = 0. \end{cases}$$

Assumption II is satisfied because the right side of (20) involves only elements of $\underset{\sim}{a}_n$; i.e. $\underset{\sim}{s}_n$ enters only as a component of $\underset{\sim}{a}_n$.

For Assumption III, we must first specify the static game Γ whose payoffs $\gamma_q(\cdot)$ are defined by (3). Let $\mu_q = E(X^q_1)$. For the remainder of this section, we write q as a subscript rather than a superscript. Here, if $g_q > 0$,

$$\begin{aligned} \gamma_q(\underset{\sim}{a}) &= K_q(\underset{\sim}{a}) + \beta_q E\{L_q[\xi(\underset{\sim}{a})]\} \\ &= -d_q + \beta_q E\left\{X_{1q}\left[d_q - g_q + h_q\left(\frac{g_q}{g_1+g_2}\right)\right]\right\} \end{aligned}$$

so

$$(21) \qquad \gamma_q(\underset{\sim}{a}) = -d_q + \beta_q \mu_q \left[d_q - g_q + h_q\left(\frac{g_q}{g_1+g_2}\right)\right].$$

Player q's available actions in Γ are A^q given by (18). Suppose $0 \leq \beta_q \mu_q \leq 1$, $q = 1,2$. Exercise 2 asks you to prove that Γ has the following EP:

$$(22)\begin{cases} d_1 = g_1 = (1/h_2)/(1/h_1 + 1/h_2)^2 \\ d_2 = g_2 = (1/h_1)/(1/h_1 + 1/h_2)^2. \end{cases}$$

The assumption $\beta_q \mu_q \leq 1$ restricts $E(X_{1q})$ to at most $1/\beta_q$. Hence, if $\beta_q \mu_q \leq 1$ for $q = 1,2$, then Assumption III is satisfied by $\underset{\sim}{a}^*$ specified by (22).

Assumption IV is $P\{\xi(\underset{\sim}{a}^*) \varepsilon S(\underset{\sim}{a}^*)\} = 1$. With (18), (19), and (21), this condition is

$$(23) \qquad P\{X_1^q \, h_q g^q/(g_1 + g_2) \geq g_q, \; q = 1,2\} = 1$$

where g_1 and g_2 are specified by (22) so

$$g_1 + g_2 = 1/(1/h_1 + 1/h_2).$$

Hence, (23) becomes

$$(24) \qquad P\{X_1^q \geq (1/h_q)/(1/h_1 + 1/h_2), \qquad q = 1,2\} = 1.$$

Therefore, if $(s_1^1, s_1^2) \geq (g^1, g^2)$ and $\beta_q \mu_q \leq 1$ for $q = 1,2$, so (24) is satisfied, then $(d_n^q, g_n^q) = (d_q, g_q)$ specified by (22) for all $n = 2,3,\ldots$, $g_1^q = g_q$ and $b_1^q = s_1^q - g_q$ comprise an EP for the SG. Using (7), the resulting expected discounted return in the infinite duration case is

$$(25) \qquad s_1^1 - [(1/h_2)/(1/h_1 + 1/h_2)^2][1 - \beta_1\mu_1 h_1(1/h_1 + 1/h_2)]/(1 - \beta_1)$$

for player 1, with a similar expression for player 2. Notice that neither the actions (22) nor their consequent expected return (25) depend on the joint probability distribution of (X_1^1, X_1^2) except through μ_1 and μ_2 i.e. the expected values of the marginal distributions.

A *symmetric game* is one in which each player has the same opportunities and rewards. In the present game, symmetry means $\beta \triangleq \beta_1 = \beta_2$, $\mu \triangleq \mu_1 = \mu_2$, and $h_1 = h_2$; then $J \triangleq J_1 = J_2$. If $h_1 = h_2 = h$, then (22) becomes

$$d^1 = g^1 = d^2 = g^2 = h/4$$

and (24) is reduced to

$$(26) \qquad P\{X_1^q \geq 1/2, \quad q = 1,2\} = 1.$$

The entire restriction on (X_1^1, X_1^2) is $\mu \leq 1/\beta$ and (26).

EXERCISES

1. Prove Theorem 2.

2. In (21), suppose $0 \leq \beta_q \mu_q \leq 1$ for $q = 1,2$. Verify that (22) specifies an EP for Γ with payoffs given by (21) and actions given by (18). (Hint: Theorem 3).

3. Suppose the Guns and Butter model is modified so H_n^q replaces h_q in (15) and we assume $(X_n^1, H_n^1, X_n^2, H_n^2)$, $n = 1,2, \ldots$, are independent and identically distributed random vectors. Let $h_q = E(H_1^q)$ and let (21) define d^q and g^q, $q = 1,2$. Prove that this modified model has the same myopic EP as the original model if (23) is replaced by

$$P\{X_1^q H_1^q \geq [(J_1 + J_2)/(J_1 J_2)](1/J_1 + 1/J_2)^{-2}, \quad q = 1,2\} = 1.$$

4. The model in this section is *stationary*, i.e. time-invariant. Suppose, instead, that S_n is the set of states in period, A_{sn}^q is player q's set of alternative actions in period n if

$$s_n = s, \quad p_n(J|s,\underset{\sim}{a}) = P\{s_{n+1} \varepsilon\, J | s_n = s, \quad \underset{\sim}{a}_n = \underset{\sim}{a}\}$$

for $J \subset S_{n+1}$ and $(s,\underset{\sim}{a}) \varepsilon\, \mathcal{G}_n = \{(s,\underset{\sim}{a}): \underset{\sim}{a} \varepsilon \times_{q \varepsilon Q} A_{sn}^q$ and $s \varepsilon S_n\}$, $r_{qn}(s,\underset{\sim}{a})$ is player q's reward in period n if $(s_n, \underset{\sim}{a}_n) = (s,\underset{\sim}{a}) \varepsilon\, \mathcal{G}_n$, and one monetary unit at the start of n+1 is worth β_{qn} at the start of period n. State and prove a version of Theorem for this *nonstationary* model. (Hint: Let

$$\gamma_{qn}(\underset{\sim}{a}) = K_{qn}(\underset{\sim}{a}) + \beta_{qn} E\{L_{q,n+1}[\xi_n(\underset{\sim}{a})]\}).$$

5. A more general version of the model in Theorem 3 is

$$\gamma_q(\underset{\sim}{a}) = \alpha_q + \omega_q a_q/(\Sigma_{i\varepsilon Q} a_i)^m - \xi_q a_q$$

with $m \varepsilon I_+$ and $m < Q$. Suppose also that $\eta = \xi_q/\omega_q$ is the same for all q. Show that

$$a_q^* = [(Q-m)/Q^{m+1}]^{1/m} \qquad \text{for all } q$$

is an EP of the static game Γ.

6. Prove that the conclusions of Theorem 3 are valid for the following model (which is different from the one in Theorem 3):

$$\gamma_q(\underset{\sim}{a}) = \alpha_q - \omega_q \Sigma_{i\neq q}\, a_i/\Sigma_{i\varepsilon Q}\, a_i - \xi_q a.$$

5.* Competitive Advertising Decisions

Section 4 of Chapter III presents an MDP model of a firm's advertising decisions. That model does not include the competitive reaction of competing firms, a feature which is added in this section. It is important to use an SG model of advertising decisions instead of an MDP model if one's competitors react quickly and their decisions greatly affect one's own demand.

This section analyzes a duopoly model, i.e. a model of two competing firms. The principal conclusions are valid in a similar model with more than two firms but the exposition would be more cumbersome. The end of section 6 combines the results in this section with those of a duopoly model in which firms make production decisions and hold inventories.

Duopoly models are sometimes useful to model competitive decisions in industries with more than two firms but where one firm is dominant. Then the two firms in the model are the dominant firm and a pseudo-"firm" which aggregates all the firms except the dominant one. Some examples of industries where the leading firm is indeed dominant are IBM in computers and General Mills in breakfast cereals.

The notation in the following model generally consists of appending "q" superscripts and subscripts to the notation in Section 4 of Chapter III. The model describes two interacting firms so $q \in Q = \{1,2\}$. We assume that the effect of firm q's advertising on its "goodwill" depreciates at a rate θ_q per time period, $0 \leq \theta_q \leq 1$. Let z_n^q denote firm q's advertising expenditure in period n so $z_n^q \theta_q^j$ is the impact of z_n^q on goodwill in period n+j. Firm q's goodwill is the aggregate impact of its advertising expenditures so a_n^q, its goodwill in period n, is

$$a_n^q = z_1^q\theta_q^{n-1} + z_2^q\theta_q^{n-2} + \ldots + z_{n-1}^q\theta + z_n;$$

(1) $$a_n^q = \Sigma_{k=0}^{n-1}\, z_{n-k}^q\theta^k = z_n^q + \theta_q a_{n-1}^q = z_n^q + s_n^q$$

where s_n^q denotes $\theta_q a_{n-1}^q$. Let $\underset{\sim}{a}_n = (a_n^1, a_n^2)$ and $\underset{\sim}{s}_n = (s_n^1, s_n^2)$.

Let D_n^q be firm q's demand in period n, measured in physical units, and let $\underset{\sim}{D}_n = (D_n^1, D_n^2)$. We assume for each n that the distribution of $\underset{\sim}{D}_n$ depends only on $\underset{\sim}{a}_n$, i.e. demand depends only on current goodwill. Let

(2) $$\mu_q(\underset{\sim}{a}) = E(D_1^q | \underset{\sim}{a}_1 = \underset{\sim}{a}), \qquad \underset{\sim}{a} \geq (0,0)$$

where we emphasize that the distribution of D_1^q may depend on *both* firms' goodwills.

Let r_q be firm q's gross profit per unit of demand, not including advertising expenditures; we assume $r_q > 0$. Then $r_q D_n^q - z_n^q$ is the gross profit in period n so

(3) $$V^q = \Sigma_{n=1}^{\infty}\, \beta_q^{n-1}(r_q D_n^q - z_n^q)$$

is firm q's sum of discounted profits; we assume $\beta_q < 1$, $q = 1,2$.

Let $s_1^q = \theta_q a_0^q$ denote the initial goodwill so $a_1^q = z_1^q + s_1^q$; then (1) is valid for all $n \varepsilon I_+$. Substitution of $z_n^q = a_n^q - \theta_q a_{n-1}^q$ in (3) yields

$$V^q = \theta_q a_0^q + \Sigma_{n=1}^{\infty}\, \beta_q^{n-1}[r_q D_n^q - (1 - \beta_q\theta_q)a_n^q].$$

Let

(4) $$\gamma_q^*(\underset{\sim}{a}) = r_q\mu_q(\underset{\sim}{a}) - (1 - \beta_q\theta_q)a^q, \qquad \underset{\sim}{a} = (a^1, a^2) \geq (0,0).$$

Let $E_{\underset{\sim}{\Pi}}$ denote an expectation with respect to probabilities induced by the firms' use of policy $\underset{\sim}{\Pi} = (\Pi_1, \Pi_2)$. Then

$$(5) \qquad v_s^q(\underset{\sim}{\Pi}) = s^q + E_{\underset{\sim}{\Pi}}[\Sigma_{n=1}^{\infty} \beta_q^{n-1} \gamma_q^*(\underset{\sim}{a}_n) | \underset{\sim}{s}_1 = \underset{\sim}{s}]$$

where $\underset{\sim}{s} = (s^1, s^2)$.

The constraint on $\underset{\sim}{a}_n$ is

$$(6) \qquad \underset{\sim}{0} \le \underset{\sim}{z}_n = \underset{\sim}{a}_n - \underset{\sim}{s}_n = (a_n^1, a_n^2) - (\theta_1 a_{n-1}^1, \theta_2 a_{n-1}^2)$$

or $\theta_q a_{n-1}^q \le a_n^q$ for each q. Note that the realized values of the demands $\underset{\sim}{D}_1, \underset{\sim}{D}_2, \ldots$ do not affect the feasibility of $\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots$.

Suppose there exists $\underset{\sim}{a}^* = (a_1^*, a_2^*)$ such that

$$(7) \qquad \gamma_1^*(\underset{\sim}{a}^*) \ge \gamma_1^*(a^1, a_2^*) \quad \text{and} \quad \gamma_2^*(\underset{\sim}{a}^*) \ge \gamma_2^*(a_1^*, a^2), \quad a^1 \ge 0,\ a^2 \ge 0.$$

Then the model satisfies Assumptions I through IV in Section 4 (Exercise 1 asks you to verify this claim) so the following result is a consequence of Theorem 4.1.

PROPOSITION 1. *If* (7) *is valid, then* $\underset{\sim}{a}_n = \underset{\sim}{a}^*$ *for all* n *is a discounted EP*[1] *with respect to* {s: $s \le \underset{\sim}{a}^*$} *and the set of all policies.*

Proposition 1 states that, if $s_1 \le \underset{\sim}{a}^*$, then $\underset{\sim}{a}_n = \underset{\sim}{a}^*$ for all n comprises an EP. From (6), the consequent advertising expenditures are

$$z_1^q = a_q^* - s_1^q \quad \text{and} \quad z_n^q = a_q^*(1 - \theta_q) \quad \text{for } n > 1.$$

---

[1]The rule $\underset{\sim}{a}_n = \underset{\sim}{a}^*$ is not a policy because $\underset{\sim}{a}_n = \underset{\sim}{a}^*$ is not feasible if $\underset{\sim}{s}_n \nleq \underset{\sim}{a}^*$. Let $\underset{\sim}{a}_n$ be arbitrary, but feasible, if $\underset{\sim}{s}_n \nleq \underset{\sim}{a}^*$. Of course, if $\underset{\sim}{s}_1 \le \underset{\sim}{a}^*$ then the rule $\underset{\sim}{a}_n = \underset{\sim}{a}^*$ causes $\underset{\sim}{s}_n \le \underset{\sim}{a}^*$ for all n.

In order to analyze the static game with payoffs in (4), it is convenient to alter the notation and write $\underset{\sim}{a} = (a_1,a_2)$ instead of (a^1,a^2). Observe that $\underset{\sim}{a}^*$ is an EP of the static game with payoffs in (4) if and only if it is an EP for the static game with payoffs $\gamma_q(\cdot) = \gamma_q^*(\cdot)/r_q$:

$$(8) \qquad \gamma_q(\underset{\sim}{a}) = \mu_q(\underset{\sim}{a}) - h_q a_q, \qquad h_q = (1 - \beta_q\theta_q)/r_q.$$

If $\gamma_q(\cdot)$ is differentiable at $(a_1,a_2) > (0,0)$, then a necessary condition for an interior EP is

$$(9) \qquad \partial\mu_q(a_1,a_2)/\partial a_q - h_q = 0, \quad q = 1,2.$$

EXAMPLE 1. Let $w_q(\underset{\sim}{a}) = a_q/(a_1+a_2)$ which is firm q's fraction of the total goodwill. Suppose D_n^q for each n and q has a marginal distribution which is a uniform distribution on the interval $[g_q, g_q + 2J_q w_q(a)]$. Then $\mu_q(\underset{\sim}{a}) = g_q + J_q w_q(\underset{\sim}{a})$ so (9) is

$$(10) \qquad J_1 a_2/(a_1 + a_2)^2 - h_1 = J_2 a_1/(a_1 + a_2)^2 - h_2 = 0.$$

Let $H_q = h_q/J_q$. From (10),

$$a_1 = H_2(a_1 + a_2)^2 \quad \text{and} \quad a_2 = H_1(a_1 + a_2)^2$$

so $a_1 + a_2 = (H_1 + H_2)^{-1}$ whose substitution in (10) yields

$$(11) \qquad a_1^* = H_2/(H_1 + H_2)^2 \quad \text{and} \quad a_2^* = H_1/(H_1 + H_2)^2.$$

In the symmetric duopoly case, $H_1 = H_2 = H$ so (11) becomes $a_1^* = a_2^* = (4H)^{-1}$. $\square$

EXERCISES

1. Suppose there are Q firms instead of only two, and $Q \geq 2$. In place of (7), if

$$\gamma_q(\underset{\sim}{a}^*) \geq \gamma_q(a_q, \underset{\sim}{a}^*_{-q}), \qquad a_q \geq 0, \qquad q = 1,\ldots,Q,$$

then Proposition 1 is still valid. Work out the details of Example 1 if $w_q(\underset{\sim}{a}) = a_q / \Sigma_{q=1}^{Q} a_q$ and D_n^q has a uniform marginal distribution on $[g_q,\ g_q + 2J_q w_q(\underset{\sim}{a})]$ for each n and q. (Hint: Theorem 4.3).

6. Dynamic Oligopoly

Sequential games constitute a reasonable framework in which to analyze many of the economic phenomena associated with imperfect competition. The preceding section presents a duopoly model which is a natural generalization of Section 4 of Chapter III. This section contains a multi-firm generalization of Sections 1 and 2 of Chapter III, i.e. a dynamic oligopoly model in which firms hold inventories from one period to the next. A firm's decisions each period are the amount to produce and the price at which it is willing to its goods. At the end of the section, we add a third decision, namely the amount to spend on advetising.

Let $\underset{\sim}{Q}$ be a set of firms and let s_n^q be firm q's inventory level at the beginning of period n. We assume that production is sufficiently rapid relative to a period's length that the quantity produced can be used to satisfy demand in the same period. Let z_n^q and D_n^q be firm q's production quantity and demand in period n, respectively. We assume $P\{D_n^q \geq 0\} = 1$ for all n and q. Suppose that excess demand is backlogged so $s_{n+1}^q = s_n^q + z_n^q - D_n^q$. Let $y_n^q = s_n^q + z_n^q$ and let $\underset{\sim}{s}_n$, $\underset{\sim}{z}_n$, $\underset{\sim}{y}_n$, and $\underset{\sim}{D}_n$ denote the vectors whose s-*th* components are s_n^q, z_n^q, y_n^q, and D_n^q, respectively. Then

$$\underset{\sim}{s}_{n+1} = \underset{\sim}{s}_n + \underset{\sim}{z}_n - \underset{\sim}{D}_n = \underset{\sim}{y}_n - \underset{\sim}{D}_n.$$

Let ρ_n^q be the price announced by firm q in period n and let $\underset{\sim}{\rho}_n$ be the vector whose s-*th* component is ρ_n^q. Let $\underset{\sim}{a}_n^q = (y_n^q, \rho_n^q)$ and let $\underset{\sim}{a}_n$ be the vector whose s-*th* component is a_n^q. We assume that the distribution of $\underset{\sim}{D}_n$, given $\underset{\sim}{a}_n = \underset{\sim}{a} = (\underset{\sim}{\rho}, \underset{\sim}{y})$, is conditionally independent of the history $\underset{\sim}{s}_1$, $\underset{\sim}{a}_1$, $\underset{\sim}{D}_1$, ..., $\underset{\sim}{s}_{n-1}$, $\underset{\sim}{a}_{n-1}$, $\underset{\sim}{D}_{n-1}$, $\underset{\sim}{s}_n$, and depends only on $\underset{\sim}{\rho}$ and, possibly, on $\underset{\sim}{y}$. Let

$$\mu_q(\underset{\sim}{a}) = E(D_n^q | \underset{\sim}{a}_n = \underset{\sim}{a}).$$

Thus the distribution of each firm's demand may be affected by the prices and quantities set by competing firms as well as by its own price and quantity.

Suppose that each firm has two kinds of costs. We assume that production cost is proportional to the amount produced; we believe that this is the major limitation of the model. Let $c_q \cdot z_n^q$ be firm q's production cost in period n. Also, suppose that firm q's salvage value of s^q is $-c_q$.

Let $g_q(\underset{\sim}{y},\underset{\sim}{\rho},\underset{\sim}{d})$ denote firm q's revenue minus its inventory-related costs in period n if $\underset{\sim}{a}_n = (\underset{\sim}{y},\underset{\sim}{\rho})$ and $\underset{\sim}{D}_n = \underset{\sim}{d}$. This representation encompasses many cases.

EXAMPLE 1. Let θ_n^q be the "raw" demand faced by firm q in period n. Suppose θ_n, the vector with s-*th* component θ_n^q, depends only on the price vector ρ_n. Let

$$D_n^q = \begin{cases} y_n^q & \text{if } y_n^q \le \theta_n^q \\ \theta_n^q + \min\{m_q \Sigma_{j\varepsilon S}(\theta_n^j - y_n^j)^+, \quad y_n^q - \theta_n^q\} & \text{if } y_n^q > \theta_n^q, \end{cases}$$

where $\Sigma_{q\varepsilon Q} m_q \le 1$. Thus the general backordering assumption encompasses the case in which excess demand is lost. The simplest such case would have $m_q=0$ in (2) for all $q\varepsilon Q$. Suppose that consumers in this industry are well informed and will pay only the lowest price set by any of the firms. Then $g_q(\underset{\sim}{y},\underset{\sim}{\rho},\underset{\sim}{d})$ might take the form

$$\text{(2)} \qquad g_q(\underset{\sim}{y},\underset{\sim}{\rho},\underset{\sim}{d}) = d_q \min\{\rho^j : j\varepsilon Q\} - h_q \cdot (y^q - d^q)^+$$
$$- b_q \cdot (d^q - y^q)^+$$

where h_q and b_q are respective unit costs of inventory and shortage. In this example, other firms' actions affect firm q via demand and via the price that traffic will bear. $\square$

Let β_q denote firm q's discount factor. Exercise 1 ask you to verify that the oligopoly model satisfies Assumptions I and II in Section 4 with (4.3) taking the form

$$(3) \qquad \gamma_q(\underset{\sim}{a}) = E[g_q(\underset{\sim}{a},\underset{\sim}{D}_1) - \beta_q c_q D_1^q | \underset{\sim}{a}_1 = \underset{\sim}{a}] - c_q(1 - \beta_q)y^q$$

where y^q is the q-*th* component of $\underset{\sim}{y}$ in $\underset{\sim}{a} = (\underset{\sim}{y},\underset{\sim}{\rho})$.

Let the constraints on $\underset{\sim}{a}_n^q = (y_n^q, \rho_n^q)$ be

$$0 \leq z_n^q = y_n^q - s_n^q \quad \text{so} \quad s_n^q \leq y_n^q \quad \text{and} \quad 0 \leq \rho_n^q,$$

so that production quantities and prices must both be nonnegative. With backordering in the model, $S = \mathbb{R}^Q$ so

$$(4) \quad A_s^q = [s^q,\infty) \times [0,\infty), \quad A^q = \bigcup_{s \in S} A_s^q = \mathbb{R}_+^2, \quad \text{and} \quad A = \underset{q \in Q}{\times} A^q = \mathbb{R}^{2Q}$$

where s^q is the q-*th* component of $\underset{\sim}{s}$. As in the general case in Section 3, let Γ denote the following static game among the firms in Q: firm q's payoffs is $\gamma_q(a)$ and its set of alternative moves is A^q.

Exercise 1 ask you to verify that, *if* Γ has an EP (equilibrium point) in pure strategies, then the model satisfies Assumption IV in Section 3. The following result is then an immediate corollary of Theorem 4.1.

THEOREM 1. *If* Γ *has an unrandomized EP* $\underset{\sim}{a}^* = (\underset{\sim}{y}^*,\underset{\sim}{\rho}^*)$, *let* α *be a single stage decision rule which specifies* $\alpha(\underset{\sim}{s}) = \underset{\sim}{a}^*$ *(with probability one) if* $\underset{\sim}{s} \leq \underset{\sim}{y}^*$ *and is arbitrary but feasible if* $\underset{\sim}{s} \not\leq \underset{\sim}{y}^*$. *Suppose* $\beta_q < 1$ *for all* $q \in Q$.

(a) α^{∞} *is a discounted EP with respect to* $[y^*,\infty)$ *and the set of all policies.*

(b) α^N *is a N-period EP, for every* $N\varepsilon I_+$, *with respect to* $[y^*,\infty)$ *and the set of all policies.*

An industry which has the structure of the oligopoly model would have the following properties if Γ has an unrandomized EP and $s_1 \leq y^*$:

(i) Each firm's price would be time-invariant. However, different firms may have different prices (the theorem does not assert that all the components of ρ^* are the same).

(ii) Each firm's maximum inventory would be time-invariant.

Suppose that the model remains myopic but no longer time-invariant (see Exercise 4.4). Then (i) and (ii) would no longer be valid. The myopic structure itself would vanish if the model included bankruptcy and other financial details.

We have assumed $P\{D_n \geq 0\} = 1$ for all n. Therefore, from (1), if $y_n = y^*$, then

$$s_{n+1} = y_n - D_n = y^* - D_n \leq y^*$$

so $y_{n+1} = y^*$ is again feasible (with probability one). As a result, the requirement in Theorem 1 that Γ has an unrandomized EP is stronger than necessary.

COROLLARY 1.1. *Suppose* Γ *has an EP* z, *possible randomized, which assigns* $P\{y = y^*\} = 1$ *for some* $y^*\varepsilon\mathbb{R}^Q$. *Let* λ *denote a single stage decision rule which specifies* $\lambda(s) = z$ *if* $s \leq y^*$ *and is arbitrary* if $s \nleq y^*$. *Suppose* $\beta_q < 1$ *for all* $q\varepsilon Q$. *Then* (a) *and* (b) *of Theorem* 1 *are valid with* λ^{∞} *and* λ^N *in place of* α^{∞} *and* α^N.

Linear Inventory and Backorder Costs

The rest of this section presents some of the cases in which the condition of the corollary is satisfied. Suppose

$$(5) \quad g_q(\underset{\sim}{y},\underset{\sim}{\rho},\underset{\sim}{d}) = \rho^d d^q - h\cdot(y^q - d^q)^+ - b\cdot(d^q - y^q)^+$$

where, unlike (2), the unit costs of inventory and backorders are assumed the same for all firms (i.e. h and b are not subscripted with q). Similarly, let $\beta_q = \beta$ and $c_q = c$ for all q. These assumptions are made merely to simplify the notation but they are reasonable in many industries.

Suppose for each q that the marginal distribution function F_q of D_1 depends on $\underset{\sim}{a}_1 = (y_1,\rho_1)$ only via the price vector $\underset{\sim}{\rho}_1$. Let

$$F_q(x|\underset{\sim}{\rho}) = P\{D_1^q \leq x|\underset{\sim}{\rho}_1 = \underset{\sim}{\rho}\}$$

so that the expected value of D_1^q can be written

$$\mu_q(\underset{\sim}{\rho}) = E[D_1^q|\underset{\sim}{a}_n = (\underset{\sim}{y},\underset{\sim}{\rho})].$$

With these simplifications, the substitution of (5) in (3) yields

$$(6') \quad \gamma_q(\underset{\sim}{a}) = (\rho^q - \beta c_q)\mu_q(\underset{\sim}{\rho}) - c_q(1-\beta)y^q - E[h(y^q - D_1^q)^+ + b(D_1^q - y^q)^+|\underset{\sim}{\rho}_1 = \underset{\sim}{\rho}].$$

Since we shall analyze several versions of Γ with payoffs as in (6'), it clarifies the exposition to write D_q for D_1^q, y_q for y^q, and ρ_q for ρ^q. Then (6') becomes

$$(6) \quad \gamma_q(\underset{\sim}{a}) = (\rho_q - \beta c)\mu_q(\underset{\sim}{\rho}) - c(1-\beta)y_q - E[h(y_q - D_q)^+ + b(D_q - y_q)^+|\underset{\sim}{\rho}_1 = \underset{\sim}{\rho}].$$

We assume for each q that $F_q(\cdot|\underset{\sim}{\rho})$ has a density function $\phi_q(\cdot|\underset{\sim}{\rho})$. Suppose Γ has an unrandomized EP $\underset{\sim}{a}^* = (\underset{\sim}{y}^*, \underset{\sim}{\rho}^*)$.

(7) $$0 = \partial\gamma_q(\underset{\sim}{a})/\partial y_q \Big|_{\underset{\sim}{a}=\underset{\sim}{a}^*}$$

is necessary. From (6),

$$\frac{\partial\gamma_q(\underset{\sim}{a})}{\partial y_q}\Bigg|_{\underset{\sim}{\rho}=\underset{\sim}{\rho}^*} = -c(1 - \beta) - hF_q(y_q|\underset{\sim}{\rho}^*) + b[1 - F_q(y_q|\underset{\sim}{\rho}^*)].$$

Therefore, (7) is equivalent to

$$F_q(y_q|\underset{\sim}{\rho}^*) = \frac{b-c(1-\beta)}{b+h} .$$

Let

$$F_q^{-1}(r|\underset{\sim}{\rho}) = \sup\{x\colon F_q(x|\underset{\sim}{\rho}) \le r\}$$

and

(8) $$v = [b - c(1 - \beta)]/[b + h]$$

where $b > c(1 - \beta)$ is assumed and satisfied in practice. Now (7) is equivalent to

(9) $$y_q = F_q^{-1}(v|\underset{\sim}{\rho}^*).$$

This equation has exactly the same form as (III.2.4) which specifies the optimal base stock level for a firm in a perfectly competitive market (or a monopolist whose price is not a decision variable).

From (9), for each j we may substitute

$$(10)\qquad y_j = F_j^{-1}(v|\underset{\sim}{\rho})$$

in (6) to obtain $\gamma_q(\cdot)$ as a function only of ρ. Let $M_q(\underset{\sim}{\rho})$ denote $\gamma_q(\underset{\sim}{a}) = \gamma_q(\underset{\sim}{y},\underset{\sim}{\rho})$ when each component of $\underset{\sim}{y}$ is replaced by $J_q(\underset{\sim}{\rho}) \triangleq F^{-1}(v|\underset{\sim}{\rho})$:

$$(11)\qquad M_q(\underset{\sim}{\rho}) = (\rho_q - \beta c)\mu_q(\underset{\sim}{\rho}) - c(1-\beta)F_q^{-1}(v|\underset{\sim}{\rho})$$

$$-h\int_0^{J_q(\underset{\sim}{\rho})} [J_q(\underset{\sim}{\rho}) - x]\phi_q(x|\underset{\sim}{\rho})dx - b\int_{J_q(\underset{\sim}{\rho})}^{\infty} [x - J_q(\underset{\sim}{\rho})]\phi_q(x|\underset{\sim}{\rho})dx.$$

Then a necessary condition for an EP is

$$(12)\qquad 0 = \partial M_q(\underset{\sim}{\rho})/\partial\rho_q\Big|_{\underset{\sim}{\rho} = \underset{\sim}{\rho}^*}$$

if $M_q(\cdot)$ is concave on $\mathbb{R}_+^Q$ and if there is any $\underset{\sim}{\rho}^*$ which satisfies (12).

In order to check (12), let

$$f_q(\underset{\sim}{\rho}) = \frac{\partial J_q(\underset{\sim}{\rho})}{\partial \rho_q}, \quad \mu_q'(\underset{\sim}{\rho}) = \frac{\partial \mu_q(\underset{\sim}{\rho})}{\partial \rho_q}, \quad \text{and} \quad \phi_q'(x|\underset{\sim}{\rho}) = \frac{\partial \phi_q(x|\underset{\sim}{\rho})}{\partial \rho_q}$$

which are assumed to exist. Then Leibnitz' rule and (10) yield

$$\text{(13)} \quad \frac{\partial M_q(\underset{\sim}{\rho})}{\partial \rho_q} = \mu_q(\underset{\sim}{\rho}) + (\rho_q - \beta c)\mu_q'(\underset{\sim}{\rho}) - c(1 - \beta)f_q(\underset{\sim}{\rho})$$

$$- h\int_0^{J_q(\underset{\sim}{\rho})} [J_q(\underset{\sim}{\rho}) - x]\,\phi_q'(x|\underset{\sim}{\rho})dx - hf_q(\underset{\sim}{\rho})F_q[J_q(\underset{\sim}{\rho})|\underset{\sim}{\rho}]$$

$$- b\int_{J_q(\underset{\sim}{\rho})}^{\infty} [x - J_q(\underset{\sim}{\rho})]\phi_q'(x|\underset{\sim}{\rho})dx + bf_q(\underset{\sim}{\rho})\{1 - F_q[J_q(\underset{\sim}{\rho})|\underset{\sim}{\rho}]\}.$$

Now $0 = \partial M_q(\underset{\sim}{\rho})/\partial \rho_q$ and $F_q[J_q(\underset{\sim}{\rho})|\underset{\sim}{\rho}] = F_q[F_q^{-1}(v|\underset{\sim}{\rho})|\underset{\sim}{\rho}] = v$ yield

$$\text{(14)} \quad \begin{aligned} \mu_q(\underset{\sim}{\rho}) + (\rho_q - \beta c)\mu_q'(\underset{\sim}{\rho}) &= b\int_{J_q(\underset{\sim}{\rho})}^{\infty} [x - J_q(\underset{\sim}{\rho})]\phi_q'(x|\underset{\sim}{\rho})dx \\ &\quad + h\int_0^{J_q(\underset{\sim}{\rho})} [J_q(\underset{\sim}{\rho}) - x]\phi_q'(x|\underset{\sim}{\rho})dx \end{aligned}$$

Direct verification of the concavity of $M_q(\cdot)$ via its Hessian matrix seems too painful without a specific assumption concerning F_q, hence f_q, μ_q, μ_q', ϕ_q, and ϕ_q'. Nevertheless, observe that neither (9) nor (12) depend on the joint distribution of all the firms' demands except through the marginal distributions. In order to simplify the analysis, the following cases have $\phi_q'(x|\underset{\sim}{\rho})$ constant with respect to x.

Uniform Marginal Distributions

EXAMPLE 2. Suppose that there are two firms which we label $q = 1$ and $q = -1$ for convenience of exposition. We assume that the marginal distribution of D_q, firm q's demand, is uniform on the interval $[0, w\rho_{-q}/\{1 + (\rho_1 + \rho_{-1})^2\}]$. These marginal distributions would result in the following cases:

(i) For each q, $D_q = w\rho_{-q}U/[1 + (\rho_1 + \rho_{-1})^2]$ where U is uniformly distributed on $[0,w]$;

(ii) For each q, $D_q = \rho_{-q}\, U$ where U is uniformly distributed on $[0, w/\{1 + (\rho_{-1} + \rho_1)^2\}]$; and

(iii) D_1 and D_{-1} are independent r.v.'s with the uniform distributions specified above.

Uniform marginal distributions yield

$$\mu_q = [w\rho_{-q}/2]/[1 + (\rho_1 + \rho_{-1})^2]$$

$$\mu_q' = w\rho_{-q}(\rho_1 + \rho_{-1})/[1 + (\rho_1 + \rho_{-1})^2]^2 \leq 0 \quad (\rho_1, \rho_{-1} \geq 0)$$

$$\frac{\partial \rho_q \mu_q}{\partial \rho_q} = \mu_q + \rho_q \mu_q' = \frac{(w-2)\rho_1\rho_{-1}(\rho_1 + \rho_{-1}) + w\rho_{-q}(1 + \rho_{-q}^2)}{2[1 + (\rho_1 + \rho_{-1})^2]^2}.$$

These relationships show that:

(a) The firm's average demand decreases as its price rises;

(b) For certain values of w(e.g. $w = 1$), firm q's expected revenue is unimodal (quasiconcave) in ρ_q;

(c) $D_q \to 0$ as $\rho_{-q} \to 0$, and $D_q \to w\rho_{-q}/(1 + \rho_{-q}^2)$ as $\rho_q \to 0$ (both with probability one).

In order to use (9) and (13) to obtain $\underset{\sim}{y}^*$ and $\underset{\sim}{\rho}^*$, we need F_q^{-1}, J_q, and ϕ_q'. By assumption,

$$\phi_q(x|\underset{\sim}{\rho}) = [1 + (\rho_1 + \rho_{-1})^2]/[w\rho_{-q}], \quad 0 \leq x \leq w\rho_{-q}/[1 + (\rho_1 + \rho_{-1})^2]$$

so

$$\phi_q'(x|\underset{\sim}{\rho}) = 2(\rho_1 + \rho_{-1})/(w\rho_{-q})$$

$$F_q^{-1}(u|\underset{\sim}{\rho}) = w\rho_{-q}U/[1 + (\rho_1 + \rho_{-1})^2], \quad 0 < u < 1,$$

(15) $$J_q(\underset{\sim}{\rho}) = F_q^{-1}(v|\underset{\sim}{\rho}) = w\rho_{-q}v/[1 + (\rho_1 + \rho_{-1})^2].$$

It is convenient to define

$$\xi = [-2\beta c + b(1-v)^2 + hv^2]/2$$

(16) $$\theta_q = 2\rho_{-q}\xi - \rho_{-q}^2 - 1.$$

Substitution in (13) yields

$$\mu_q - (\rho_q - \beta c)(\rho_1 + \rho_{-1})(\mu_q/2)/[1 + (\rho_1 + \rho_{-1})^2]$$

$$= [2(\rho_1 + \rho_{-1})/(w\rho_{-q})\{b\int_{J_q(\underset{\sim}{\rho})}^{\mu_q/2} [x - J_q(\underset{\sim}{\rho})]dx + h\int_0^{J_q(\underset{\sim}{\rho})} [J_q(\underset{\sim}{\rho}) - x]dx\}$$

$$= 4\mu_q^2(\rho_1 + \rho_{-1})[b(1-v)^2 + hv^2]/(w\rho_{-q}).$$

Therefore,

$$1 + (\rho_1 + \rho_{-1})^2 = 2(\rho_1 + \rho_{-1})^2(\rho_q + \xi),$$

$$\rho_q^2 + 2\xi\rho_q + \theta_q = 0,$$

and

(17) $$\rho_q = -\xi + \sqrt{\xi^2 - \theta_q}$$

if $\xi^2 \geq \theta_q$, i.e. if $\xi^2 - 2\rho_{-q}\xi + \rho_{-q}^2 + 1 \geq 0$. We shall have to verify

(18) $$\rho_{-q}^2 - 2\xi\rho_{-q} + \xi^2 + 1 \geq 0$$

when $\underset{\sim}{\rho} = \underset{\sim}{\rho}^*$.

The substitution of (16) in (17) yields

$$(\rho_q + \xi)^2 = \xi^2 + 1 + \rho_{-q}^2 - 2\xi\rho_{-q}$$

so

(19) $$2\xi(\rho_1 + \rho_{-1}) - 1 = \rho_{-q}^2 - \rho_q^2.$$

This is valid for q=1 and q= -1 so $\rho_1 = \rho_{-1}$ reduces (19) to

(20) $$\rho_1^* = \rho_{-1}^* = (4\xi)^{-1}.$$

The substitution of (20) in (15) yields the optimal "order-up-to" quantity

$$y_1^* = y_2^* = wv\xi/(1 + 4\xi^2).$$

In order to verify (17), substitute (19) to obtain

$$\rho_{-q}^2 - 2\xi\rho_{-q} + \xi^2 + 1 = (16\xi^2)^{-1} - 1/2 + \xi^2 + 1$$

$$= (16\xi^2)^{-1} + \xi^2 + 1/2 \geq 0. \quad \square$$

EXERCISES

1. (a) Verify that the oligopoly model satisfies Assumptions I and II in Section 4 with (4.3) taking the form (4).

   (b) Suppose that the static game Γ (described below (5)) has an unrandomized EP $\underset{\sim}{a}^* = (\underset{\sim}{y}^*, \underset{\sim}{\rho}^*)$. Verify that the model satisfies Assumption IV in Section 4.

2. Consider the non-symmetric version of Example 2 in which ξ_1 and ξ_{-1} may differ. Prove that EP prices satisfy

$$\rho_{-1}^*/\rho_1^* = \{\xi_1 - \xi_2 + [(\xi_1 - \xi_2)^2 - 4(\xi_2 - \xi_1 - 1)]^{\frac{1}{2}}\}/2.$$

7.* Ordinal Sequential Games

Section 3 presents sufficient conditions for an SG (sequential game) to possess an EP (equilibrium point). The proofs there depend on the following observation: if all players but one use Markov policies then the remaining player faces an MDP (Markov decision process). That fact permits us to invoke existence theorems for MDP optima but it does not lead to the straightforward application of MDP algorithms to the computation of SG EP's. As a result, the section which concerns EP algorithms, Section 8, is a collection of largely unrelated special cases.

This section has two purposes. First, we show that SG's with discounted payoffs can be imbedded in the ordinal framework of Sections 3 and 4 of Chapter IV and Section 4 of Chapter V. The second purpose is to justify a version of the policy improvement algorithm for EP's of discounted SG's. Recall from Sections 3 and 4 of Chapter IV that it is convenient to obtain results for deterministic decision processes and then to extend them to stochastic decision processes. The following deterministic SG model can be similarly extended.

Let S be a nonempty set of *states*, Q a nonempty set of *players*, A_s^q a nonempty set of *actions* for player q in states s,

$$A_s = \times_{q \in Q} A_s^q,$$

(1) $$\mathcal{S} = \{(s,a) : a \in A_s,\ s \in S\}, \qquad \Delta = \times_{s \in S} A_s, \quad \text{and} \quad Y = \times_{n=1}^{\infty} \Delta.$$

Let M be a mapping from $\mathcal{S}$ to S which determines successive states via $s_{n+1} = M(s_n, a_n)$. The set of all *posterities* with initial state s is

(2) $$\Phi_s = \{(s_1,a_1,s_2,a_2,\ldots): s_1 = s \text{ and } (s_n,a_n)\varepsilon \mathcal{G} \quad \text{and}$$

$$s_{n+1} = M(s_n,a_n) \text{ for all } n\}.$$

Definitions (1) and (2) are formally identical to (V.4.11) and (V.4.12) which specify an ordinal framework for multiple criteria MDP's. In Section 4 of Chapter V, $\mathcal{Q}$ denotes a set of criteria; here, the elements of $\mathcal{Q}$ are players. With that difference in interpretation, the following results are immediately applicable to SG's: Proposition V.4.1, Lemma V.4.1, Theorems V.4.3 and V.4.4, and Corollary V.4.4.1. The following exposition is based partly on those results and uses the same definitions and notation. Therefore, we urge the reader to review the portions of Section 4 in Chapter V labeled "Pareto Optimal Policies" and "Solutions in Stationary Policies."

DEFINITION 1. *A* coalition *is a nonempty subset of players. The set* Ω *of* latent coalitions *is a nonempty collection of coalitions.*

For each $s\varepsilon S$ and $\omega\varepsilon\Omega$, let $\theta_s^\omega \subset \Phi_s \times \Phi_s$ indicate the preferences of coalition ω among posterities when the initial state is s. We interpret $(\tau,\tau')\varepsilon\theta_s^\omega$ as "coalition ω regards posterity τ as being at least as desireable as posterity τ' if s_1=s."

As in Section 2, let Π_q label the q-*th* player's portion of $\Pi\varepsilon Y$ where

$$Y = \underset{n=1}{\overset{\infty}{\times}} \underset{s\varepsilon S}{\times} \underset{q\varepsilon \mathcal{Q}}{\times} A_s^q.$$

The portion of Π due to all the other players is labeled Π_{-q} and we sometimes write $\Pi = (\Pi_q,\Pi_{-q})$ instead of $\Pi = (\delta_1,\delta_2,\ldots)$ where the q-*th* component of $a_n = \delta_n(s_n)$ is player q's action in period n. For each $s\varepsilon S$ and $\Pi = (\delta_1,\delta_2,\ldots)\varepsilon Y$, let $\tau_s(\Pi)$

denote the posterity generated by the Markov policy Π from the initial state s:

$$\tau_s(\Pi) = (s,\delta_1(s),\ M[s,\delta_1(s)],\ \ \delta_2\{M[s,\delta_1(s)]\},\ \ldots.).$$

DEFINITION 2. *Let* Π *and* ξ *be Markov policies.*

(3) $\Pi \underset{\sim p}{>} \xi \iff$ *either* $[\tau_s(\Pi),\tau_s(\xi)]\varepsilon\theta_s^\omega$ *for all* $s\varepsilon S$ *and* $\omega\varepsilon\Omega$,

or there are s *and* j *in* S *and* ω *and* u *in* Ω *such that*

$$[\tau_s(\Pi),\tau_s(\xi)]\notin\theta_s^\omega \ \ \textit{and}\ \ [\tau_j(\xi),\tau_j(\Pi)]\notin\theta_j^u.$$

If $\underset{\sim}{\Omega} \subset \Omega$, *then*

(4) $\Pi \underset{\sim e}{>} \xi \iff [\tau_s(\Pi),\tau_s(\xi_q,\Pi_{-q})]\varepsilon\theta_s^q$ *for all* $s\varepsilon S$ *and* $q\varepsilon\underset{\sim}{\Omega}$

We repeat Definition V.4.2.

DEFINITION 3. *If* D *is a nonempty set and* $B \subset D \times D$, *then* $b\varepsilon D$ *is* B-maximal *if* $(b,c)\varepsilon B$ *for all* $c\varepsilon D$.

These definitions lead to specifications of the *core* and an EP in terms of $\underset{\sim p}{>}$ and $\underset{\sim e}{>}$, respectively.

DEFINITION 4. *A Markov policy* Π *is an* equilibrium point (EP) $\iff$ Π *is* $\underset{\sim e}{>}$*-optimal.* *A Markov policy* Π *is in the* core $\iff$ Π *is* $\underset{\sim p}{>}$*-optimal.* *If* $\Omega = \underset{\sim}{\Omega}$ *then policies in the core are called* Pareto optima (PO).

Neither $\underset{\sim e}{>}$ nor $\underset{\sim p}{>}$ are necessarily transitive and, as we observe in Section 4 of Chapter V, intransitivity prevents the straightforward use of the arguments in Sections 3 and 4 of Chapter IV.

EXAMPLE 1. Consider the following static bimatrix game (i.e. two-player noncooperative game) in which each player has two alternative actions.

		Player 2's Action	
		1	2
Player	1	0,1	3,0
1's Action	2	2,2	0,0

The entries in each cell are the rewards garnered by players 1 and 2, respectively. This is the same array of numbers as in Example V.4.4. Let (i,j) denote the policy in which player 1 takes action i and player 2 takes action j. Then (1,1) $\gtrsim_p$ (1,2) because player 2's reward is 1 at (1,1) but only 0 at (1,2). Also, (1,2) $\gtrsim_p$ (2,1) because player 1's reward is 3 at (1,2) but only 2 at (2,1). However, (1,1) $\not\gtrsim_p$ (2,1) because both players' rewards are lower at (1,1) than at (2,1). Therefore, $\gtrsim_p$ is *not* transitive. $\square$

EXAMPLE 2. Consider the following bimatrix game.

		Player 2's Action		
		1	2	3
Player	1	0,0	0,-1	0,1
1's	2	-1,0	0,0	0,-1
Action	3	1,0	-1,0	0,0

Using the same notation as in Example 1, (1,1) $\gtrsim_e$ (2,2) because player 1 is no better off at (2,1) than at (1,1), and player 2 is no better off at (1,2) than at (1,1). Also, (2,2) $\gtrsim_e$ (3,3) because player 1 is no better off at (3,2) than at (2,2) and player 2 is no better off at (2,3) than at (2,2). However, (1,1) $\not\gtrsim_e$ (3,3) because player 1 is better off at (3,1) than at (1,1) and player 2 is better off at (1,3) than at (1,1). Therefore, $\underset{\sim}{>}_e$ is *not* transitive. $\square$

Partial Resolution of Intransitivity

We repeat Definition V.4.5.

DEFINITION 5. *Let* D *be a nonempty set and* $B \subset D \times D$. *An* inconsistency cycle connects x and y under B *if there is a finite sequence* $x_1,\ldots,x_n$ *such that* $x_1 = x_n = x$, $x_k = y$ for some $1 < k < n$, $(x_i,x_{i+1}) \in B$ for all $i < n$, and $(x_{i+1},x_i) \notin B$ for some i. *The* completion of a binary relation (B,D), *written* (B',D), *is*

$$B' = B \cup \{(x,y):\ (x,y) \in D \times D,\ (x,y) \notin B,\ \textit{and}\ (y,x) \notin B\}.$$

The transitive completion of a binary relation (B,D), *written* (B_c,D), *is*

$$B_c = B' \cup \{(x,y):\ \text{there is an inconsistency cycle which connects x and y under } B'\}.$$

See Example V.4.5 for the completion and the transitive completion of $\underset{\sim}{>}_p$ in the bimatrix game of Example 1 in this section.

Let $(\geq_c,Y)$ denote the transitive completion of $(\underset{\sim}{>}_p,Y)$. The following restatement of Corollary V.4.4.1 uses terminology in Definitions V.4.6 and V.4.7.

THEOREM 1. *If S is reachable and* $\{\theta_s^\omega\colon s \in S$ *and* $\omega \in \Omega\}$ *has consistent choice and is continuous then* $\underset{\sim}{>}_p$ *and* $\geq_c$ *have the following properties:*

(5) $T_\delta \Pi \geq_c \Pi \Rightarrow \delta^\infty \geq_c \Pi$;

(6) *If* $\Pi \in Y$ *is* $\underset{\sim}{>}_p$*-maximal then* Π *is* $\geq_c$*-maximal and there is a* $\geq_c$*-maximal* δ^∞;

(7) *If* $\Pi \in Y$ *is* $\underset{\sim}{>}_p$*-maximal then* $\Pi \geq_c T_\delta \Pi$ *for all* $\delta \in \Delta$;

(8) *If* $\#\mathcal{S} < \infty$ *then there is a* $\geq_c$*-maximal* δ^∞;

(9) *For every* $\gamma \in \Delta$, *either* γ^∞ *is* $\geq_c$*-maximal or there is another* $\delta \in \Delta$ *such that* $\delta^\infty \geq_c \gamma^\infty$.

In particular (6) asserts that, if the core is nonempty then there is a stationary policy which is in the core defined with the transitive completion of $\underset{\sim}{>}_p$. From (8), a finite SG necessarily has a stationary policy in its (nonempty) core defined with the transitive completion of $\underset{\sim}{>}_p$.

Recall that $\underset{\sim}{>}_e$ is the binary relation underlying an EP. Let $(\geq, Y)$ denote the transitive completion of $(\underset{\sim}{>}_e, Y)$. Exercise 1 asks you to prove the following result.

THEOREM 2. *If* $Q \subset \Omega$ *and* $\{\theta_s^q\colon s \varepsilon S$ *and* $q \varepsilon Q\}$ *is continuous and has consistent choice, then* (5) *through* (9) *are valid with* $\underset{\sim}{>}_e$ *and* $\geq$ *in place of* $\underset{\sim}{>}_p$ *and* $\geq_c$*, respectively.*

EXERCISES

1. Prove Theorem 2.

2. Suppose the state-to-state mapping M is nonstationary. That is, for each t, let S_t, $A_s^q(t)$, and M_t depend on $t = 1,2,\ldots$ and suppose $M_t(s,a)\varepsilon S_{t+1}$ for all $s\varepsilon S_t$ and $a\varepsilon \times_{q\varepsilon Q} A_s^q(t)$. Let $s_{t+1} = M_t(s_t,a_t)$. Say that $s\varepsilon S_{t+1}$ is *reachable* if there exists $v\varepsilon S_t$ and $a\varepsilon \times_{q\varepsilon Q} A_s^q(t)$ such that $s = M_t(v,a)$; and S_{t+1} is reachable if all $s\varepsilon S_{t+1}$ are reachable. (a) Prove that Theorem 2 remains valid. (b) Prove that Theorem 1 remains valid if S_t is reachable for all $t = 2,3,\ldots$.

TRUTH-TELLING, DOMINANT STRATEGIES AND ITERATIVE GROVES MECHANISMS*

Susan I. Cohen*
College of Management
Georgia Institute of Technology
Atlanta, Georgia 30332

*This research was partially supported by National Science Foundation Grant SOC 78-05770. The author would like to thank an annonymous referee from another journal whose request for a counter example lead to the results contained herein.

Truth-Telling, Dominant Strategies and Iterative Groves Mechanisms

ABSTRACT

This paper demonstrates that if a price-decomposition algorithm is used to solve a decentralized resource allocation problem, where rewards are those of a Groves methanism, truth-telling remains a dominant strategy equilibrium. It has been shown previously that for a general class of non-linear programming algorithms, truth-telling is a weakly dominant Nash equilibrium, but not necessarily a dominant strategy equilibrium.

Running head:

Truth-Telling and Dominant Strategies

I. INTRODUCTION

A problem that has received some attention of late is how a decentralized organization might construct an evaluation scheme for its members so that the members' own self interests coincide with the interests of the organization as a whole [1,2,4,5]. A "Groves scheme" [4] is one such mechanism. In a Groves scheme, the members of the organization communicate with a central decision maker who uses information from all the members to choose the optimal action for the organization as a whole. The messages sent by the members are in the form of functions; examples are, profit contribution functions or willingness-to-pay functions. Given this type of communication and the method of evaluation defined in a Groves mechanism, truth-telling can be shown to be a dominant strategy for the members of the organization.

From an implementation perspective, requiring communicable knowledge of an entire function is a very restrictive assumption. If messages are to be in Euclidean space rather than function space, iterative procedures must be used to solve the overall organization's problem. In an iterative procedure, the members of the organization must only communicate certain well-chosen points on the graph of the appropriate functions. It was shown in [2] that for a particular (wide) class of mathematical programming algorithms used to solve the organization's problem iteratively, truth-telling is no longer a dominant strategy for the members. However, it was shown that truth-telling is a Nash equilibrium that weakly dominates all other Nash equilibria for all members of the organization.

In this paper, we show that truth-telling is a dominant strategy when the iterative procedure used is a price-decomposition algorithm. Examples of well-known price-decomposition algorithms are the Dantzig-Wolfe Decomposition Algorithm [3] and a resource allocation procedure due to Malinvaud [1,6].

II. THE MODEL

The model used here is one developed in [2], modified for the specific class of algorithms considered. To briefly summarize [2], the basic model is of a multidivisional firm (for expository purposes only) with a corporate center (the Center) and N divisions, indexed by $i = 0,1,\ldots,N$, respectively. Profit contribution functions for the Center and the divisions are concave functions of a vector of resource use:

$$\begin{aligned} &x_i \in \mathbb{R}^L && \text{resource use of } i = 0,1,\ldots,N \\ &\pi_i\colon \mathbb{R}^L \to \mathbb{R} && \text{profit contribution function of } i = 0,1,\ldots,N \end{aligned} \tag{1}$$

The "technological coefficients" for the Center and the divisions are contained in M x L dimensional matrices α_i; resource availability is contained in an M- dimensional vector K and the set of feasible x from the centers' point of view is denoted by the convex set X.[1,2]

The Center wants to find that vector of resource use $x^* = (x_0^*, x_1^*, \ldots, x_N) \in \mathbb{R}^{L(N+1)}$ that solves the following non-linear constrained maximization problem:[3]

$$\begin{gathered} \underset{x\in X}{\text{MAX}} \sum_{i=0}^{N} \pi(x_i) \\ \sum_{i=0}^{N} \alpha_i x_i \leq K \end{gathered} \tag{2}$$

It is assumed that the Center knows only that the $\pi_i(\cdot)$ functions are concave, but not their functional form for $i = 1,\ldots,N$. The Center knows its <u>own</u> $\pi_0(\cdot)$ function in its entirety, as well as all the technological coefficients a_i; it also knows the vector of resource availabilities K and the set X. Each division has knowledge only of its own profit contribution function and technological coefficients, but not necessarily of the functions and coefficients of the other divisions. The Center and all the divisions know the general form of the problem in (2).

In [2] an <u>iterative control mechanism</u> was defined as a specification of iterative communication rules, a decision rule and enforcement rules. The particular algorithm used to solve (2) implicitly defines the communication and decision rules. The communication rules contain a message space from which the divisions choose their messages to the Center, and a message rule that specifies how the Center chooses its replies to the divisions. For the price-decomposition algorithms considered here, an iterative control mechanism C is defined as follows:

$$C \equiv (M, \{p^t(\cdot)\}_{t=1}^{\infty}, x(\cdot), \langle E_i(\cdot)\rangle_{i=1}^{N}) \tag{3}$$

where

(a) $M = \mathbb{R}^{L+1}$ is the message space of the divisions.

Division i's message at iteration t is $m_{it} = (m_{it}^1, m_{it}^2)$, where the Center interprets $m_{it}^1 \in \mathbb{R}^L$ as resource use x_i and m_{it}^2 is interpreted as profit contribution at that resource use (i.e., $\pi_i(x_i)$)

$m_i^t \equiv (m_{i1},\ldots,m_{it})$

$m^t \equiv (m_1^t,\ldots,m_N^t)$

$m_t \equiv (m_{1t},\ldots,m_{Nt})$

(b) $p^t: \mathcal{M}^{N(t-1)} \to \mathbb{R}^M$ is the message rule of the Center at iteration t, i.e., $p^t(m^{t-1})$ is the message from the Center to the divisions at iteration $t \geq 2$; p^1 is given a priori. The message $p^t(m^{t-1})$ is the vector of optimal dual multipliers associated with the first M constraints of the following constrained maximization problem:

$$\max_{\{\lambda_{i\tau}\},\, x_0} \sum_{i=1}^{N} \sum_{\tau=1}^{t-1} \lambda_{i\tau} m_{i\tau}^2 + \tau_0(x_0)$$

$$\text{s.t.} \sum_{i=1}^{N} \sum_{\tau=1}^{t-1} \alpha_i \lambda_{i\tau} m_{i\tau}^1 + \alpha_0 x_0 \leq K$$

$$\left(x_0, \sum_{\tau=1}^{t-1} \lambda_{1\tau} m_{1\tau}^1, \ldots, \sum_{\tau=1}^{t-1} \lambda_{N\tau} m_{n\tau}^1\right) \varepsilon X$$

$$\sum_{\tau=0}^{t-1} \lambda_{i\tau} = 1 \qquad i = 1,\ldots,N$$

$$\lambda_{i\tau} \geq 0 \qquad \begin{array}{l} i = 1,\ldots,N \\ \tau=0,1,\ldots,t-1 \end{array}$$

Let $(\{\lambda_{i\tau}(m^{t-1})\},\ x_0(m^{t-1}))$ denote the optimal solution to the above.

(c) $x\colon \mathcal{M}^{\infty} \to \mathbb{R}^{L(N+1)}$ is the Center's decision rule.

$$x(m^{\infty}) = (x_0(m^{\infty}),\ x_1(m^{\infty}),\ldots,x_N(m^{\infty})) \equiv$$

$$\lim_{t\to\infty} \left(x_0(m^{t-1}),\ \sum_{\tau=1}^{t-1} \lambda_{1\tau}(m^{t-1})m_{1\tau}^{1},\ \ldots,\ \sum_{\tau=1}^{t-1} \lambda_{N\tau}(m^{t-1})m_{N\tau}^{1}\right)$$

(d) $E_i\colon \mathbb{R} \times \mathcal{M}^{\infty} \to \mathbb{R}$ is the evaluation measure for division i. Let π_i denote division i's actual (observable aposteriori) realized profit contribution. Then,

$$E_i(\pi_i, m^{\infty}) \equiv \begin{cases} \pi_i + \sum_{j\neq i} \lim_{t\to\infty} \sum_{\tau=1}^{t-1} \lambda_{j\tau}(m^{t-1})m_{j\tau}^{1} + \pi_0(x_0(m^{\infty})) & \text{if } m_i^{\infty} \text{ is consistent with a function}^{4} \\ -\infty & \text{otherwise} \end{cases}$$

At each iteration of the communication process the divisions must decide how to respond to the messages from the Center, i.e., choose an $m_{it} \in \mathcal{M}$. A strategy for division i is a function u_i from $\mathbb{R}^{M}$ to $\mathbb{R}^{L+1}$ where for each t, $u_i(p^t) \equiv (m_{it}^{1}, m_{it}^{2})$. Since p^t depends upon m^{t-1}, we can express the divisions' messages recursively as functions of $u = (u_1,\ldots,u_N)$. The divisions' messages can then be expressed as $m_{it}(u) = (m_{it}^{1}(u),\ m_{it}^{2}(u))$. The set of possible divisional strategies is denoted by U.

In order to define what is meant by truth-telling, the principle behind a price decomposition algorithm must be clarified.[5] At each iteration, the divisions solve a maximization problem using a vector of dual variables generated by the Center to "price-out" the vectors $\alpha_i x_i$.

If the divisions were following the rules of the algorithm, at each iteration they would calculate m_{it} as follows:

$$m_{it}^{2} - p^{t}\alpha_i m_{it}^{1} = \max_{x_i} \pi_i(x_i) - p^{t}\alpha_i x_i \tag{4}$$

A <u>truthful strategy</u> u_i associated with any function $F_i(\cdot)$ is defined to be any $u_i \in \Psi[F_i]$ where $m_i(u)$ satisfies

$$m_{it}^{2}(u) - p^{t}\alpha_i m_{it}^{1}(u) \quad = \quad \max_{x_i} F_i(x_i) - p^{t}\alpha_i x_i \tag{5}$$

$$\text{for all } t, \text{ for any } u\backslash u_i \in U^{N-1}$$ [6]

Thus, division i is playing a truthful strategy if it follows the rules of the price decomposition algorithm no matter what the other N-1 divisions do. Given any sequence of messages $\hat{m}_i^{\infty}$ one can find a function $\hat{F}_i(\cdot)$ such that $\hat{u}_i \in \Psi[F_i]$ and $m_{it}(u/\hat{u}_i) = \hat{m}_{it}$ for all t.[7]

It is now possible to define payoff functions for the divisions which depend on the strategy vector u.

$$W_i[u] \equiv E_i[\pi_i, m^{\infty}(u)] \tag{6}$$

The set U and the functions $W_i[\cdot]$ define an N-person non-cooperative game in normal form.

III. RESULTS

The main result of this paper is that $u_i^{*} \in \Psi[\pi_i]$ is a dominant strategy for division i, i.e., truth-telling by all divisions constitutes a dominant strategy equilibrium. In order to prove this result, three preliminary results must be shown. First, it must be shown that the maximizer of a function is also the maximizer of the concave hull of the function and that the two functions are equal at this maximizing value. It is then demonstrated that the messages generated by the strategies associated with a function and the function's concave hull are identical. The third preliminary result is that the payoff to any division is the same whether the other N-1 divisions play strategies associated with particular functions or the functions' concave hulls.

The proofs of these results and the main theorem appear in the Appendix.

LEMMA 1

Let $y^* < \infty$ solve

$$\max_y f(y) \quad , \quad \text{where } f: \mathbb{R}^s \to \mathbb{R}$$

(i) Then y^* solves

$$\max_y \bar{f}(y) \quad , \quad \text{where } \bar{f}: \mathbb{R}^s \to \mathbb{R} \text{ and}$$

(ii) $f(y^*) = \bar{f}(y^*)$,

where $\bar{f}(\cdot)$ is the concave hull of $f(\cdot)$, i.e.

$$\bar{f}(y) = \sup \{\lambda_1 f(y^1)+\ldots+\lambda_s f(y^s)\}$$

$$\lambda_j \geq 0$$

$$\sum_{j=1}^{s} \lambda_j = 1$$

$$\sum_{j=1}^{s} \lambda_j y^j = y$$

LEMMA 2

Let $u_j \in \Psi[H_j]$ and $\bar{u}_j \in \Psi[\bar{H}_j]$ for all $j \neq i$. Then for any $u_i \in U$,

$m_{kt}(\bar{u}/u_i) = m_{kt}(u)$ for all $k = 1,\ldots,N$ and t.

LEMMA 3

Let $u_j \in \Psi[H_j]$ and $\overline{u}_j \in \Psi[\overline{H}_j]$ for all $j \neq i$. Then for any $u_i \in U$,

$W_i[u/u_i] = W_i[\overline{u}/u_i]$

The main result of this paper can now be stated.

THEOREM 1

Let $u_i^* \in \Psi[\pi_i]$. Then

$W_i[u/u_i^*] \geq W_i[u] \;\forall\; u_i \in U, \quad$ for any $u \backslash u_i \in U^{N-1}$

IV. SUMMARY

It has been shown that truth-telling retains its dominant strategy property when a Groves mechanism is made iterative using a price-decomposition algorithm. The usefulness of this result is that it lends further credence to the use of a price decomposition algorithm as a decentralized resource allocation procedure.

APPENDIX

Proof of Lemma 1

(i) By assumption

$$f(y^*) \geq f(y) \;\forall\; y \in \mathbb{R}^s$$

By definition of $\bar{f}(\cdot)$,

$$\bar{f}(y^*) \geq f(y^*)$$

Therefore,

$$\bar{f}(y^*) \geq f(y) \;\forall\; y \in \mathbb{R}^s$$

Thus, $\forall\; \tilde{y} \in \mathbb{R}^s$

$$\lambda_j \bar{f}(y^*) \geq \lambda_j f(y^j) \;\forall\; \lambda_j \geq 0 \text{ and } y^j \ni \sum_{j=1}^{s} \lambda_j y^j = \tilde{y} \text{ and } \sum_{j=1}^{s} \lambda_j = 1$$

Therefore, summing over j,

$$\bar{f}(y^*) \geq \sum_{j=1}^{s} \lambda_j f(y^j) \;\forall\; \lambda_j \geq 0 \text{ and } y^j \ni \sum_{j=1}^{s} \lambda_j y^j = \tilde{y} \text{ and } \sum_{j=1}^{s} \lambda_j = 1$$

Taking the supremum of both sides over all $\lambda_j \geq 0$ and $y^j \ni \sum_{j=1}^{s} \lambda_j y^j = \tilde{y}$ and $\sum_{j=1}^{s} \lambda_j = 1$, we get

$$\bar{f}(y^*) \geq \bar{f}(\tilde{y}) \;\forall\; \tilde{y} \in \mathbb{R}^s$$

(ii) Using the results in (i), the definition of $\bar{f}(y^*)$, and $f(y^*) \geq f(y)$ $\forall\; y \in \mathbb{R}^s$, we get $\bar{f}(y^*) = f(y^*)$. ∎

Proof of Lemma 2

By definition of u_j, for any $u\backslash u_j \in U^{N-1}$

$$m^2_{jt}(u/u_j) - p^t(m^{t-1}(u/u_j))\,\alpha_j m^1_{jt} = \underset{x_j}{\text{MAX}}\; H_j(x_j) - p^t(m^{t-1}(u/u_j))\cdot\alpha_j x_j$$

In order to use Lemma 1, it must be shown that

$$\overline{H_j(x_j) - y\alpha_j x_j} = \overline{H}_j(x_j) - y\alpha_j x_j$$

i.e., the concave hull of $[H_j(\cdot)$ minus $y\alpha_j x_j]$ equals the concave hull of $H_j(\cdot)$ minus $y\alpha_j x_j$.

$$\overline{H_j(x_j) - y\alpha_j x_j} \equiv \sup_{\substack{\lambda_s \geq 0 \\ \sum_{s=1}^{L} \lambda_s = 1 \\ \sum_{s=1}^{L} \lambda_s x_j^s = x_j}} \{\lambda_1(H_j(x_j^1) - y\alpha_j x_j^1) + \ldots + \lambda_L(H_j(x_j^L) - y\alpha_j x_j^L)\}$$

For all $\lambda_s \geq 0$ and $x_j^s \ni \sum_{s=1}^{L} \lambda_s = 1$, $\sum_{s=1}^{L} \lambda_s x_j^s = x_j$, we know that $\sum_{s=1}^{L} y\alpha_j \lambda_s x_j^s = y\alpha_j x_j$.

Therefore,

$$\overline{H_j(x_j) - y\alpha_j x_j} = \sup_{\substack{\lambda_s \geq 0 \\ \sum_{s=1}^{L} \lambda_s = 1 \\ \sum_{s=1}^{L} \lambda_s x_j^s = x_j}} \{\lambda_1 H_j(x_j^1) + \ldots + \lambda_L H_j(x_j^L)\} - y\alpha_j x_j = \overline{H}_j(x_j) - y\alpha_j x_j$$

The proof can now proceed by induction:

At $t = 1$, p^1 is given, so it does not depend upon u; therefore m_{j1} depends only on p^1:

$$m_{j1}^2(u/u_i) - p^1\alpha_j m_{j1}^1(u/u_i) \equiv \underset{x_j}{\text{MAX}}\ H_j(x_j) - p^1\alpha_j x_j$$

$$m_{j1}^2(\overline{u}/u_i) - p^1\alpha_j m_{j1}^1(\overline{u}/u_i) \equiv \underset{x_j}{\text{MAX}}\ \overline{H}_j(x_j) - p^1\alpha_j x_j$$

Using Lemma 1

$$m_{j1}^{1}(u/u_i) = m_{j1}^{1}(\bar{u}/u_i) \text{ and } m_{j1}^{2}(u/u_i) = m_{j1}^{2}(\bar{u}/u_i)$$

Assume

$$m_{j\tau}^{1}(u/u_i) = m_{j\tau}^{1}(\bar{u}/u_i) \text{ and } m_{j\tau}^{2}(u/u_i) = m_{j\tau}^{2}(\bar{u}/u_i)$$

for all $\tau \leq t-1$.

Then $p^t(m^{t-1}(\bar{u}/u_i)) = p^t(m^{t-1}(u/u_i))$.

Again using Lemma 1 and the definitions of u_j and $\bar{u}_j$,

$$m_{jt}^{1}(u/u_i) = m_{jt}^{1}(\bar{u}/u_i) \text{ and } m_{j\tau}^{2}(u/u_i) = m_{j\tau}^{2}(\bar{u}/u_i)$$

Since $m_{jt}(u/u_i) = m_{jt}(\bar{u}/u_i) \quad \forall\ j \neq i$, it follows immediately that

$m_{it}(\bar{u}/u_i) = m_{it}(u/u_i)$ for all t.

Therefore,

$$m_{kt}(u/u_i) = m_{kt}(\bar{u}/u_i) \quad \forall\ k = 1,\ldots,N \text{ and } \forall\ t. \ \blacksquare$$

<u>Proof of Lemma 3</u>

By definition

$$W_i[u] = \begin{cases} \pi_i + \sum\limits_{j\neq 1} \lim\limits_{t\to\infty} \sum\limits_{\tau=1}^{t} \lambda_{j\tau}(m^{t-1}(u)) \cdot m_{j\tau}^{1}(u) + \pi_0(x_0(m^{\infty}(u)) & \text{if } m_i^{\infty} \text{ is consistent with a function} \\ -\infty & \text{otherwise} \end{cases}$$

Since $m_i^{\infty}(u) = m_i^{\infty}(\bar{u}/u_i)$, if $W_i[u] = -\infty$ it is, then immediate that $W_i[\bar{u}/u_i] = -\infty$.

Suppose $m_i^{\infty}(u)$ is consistent with a function. From Lemma 2, $m^{\infty}(u) = m^{\infty}(\bar{u}/u_i) \quad \forall\ k=1,\ldots,M$. Therefore $\lambda_{j\tau}(m^{t-1}(u)) = \lambda_{j\tau}(m^{t-1}(\bar{u}/u_i))$ (from the definition of the $\lambda_{j\tau}(\cdot)$) for all t.

The desired result follows immediately. $\blacksquare$

Proof of Theorem 1

In [2], the following result was shown:

If $\hat{u}_j \in \Psi[\hat{H}_j]$, where $\hat{H}_j$ is concave, then

$W_i[\hat{u}/u_i^*] \geq W_i[\hat{u}/u_i] \quad \forall\, u_i \in U$ (i.e., if the other N-1 divisions play strategies associated with concave functions. then i's optimal strategy choice is the truthful strategy).

The current proof will proceed by contradiction:

Assume $\exists\, \tilde{u}_i \in U$ such that

$$W_i[u/\tilde{u}_i] > W_i[u/u_i^*]$$

From Lemma 3,

$$W_i[u/\tilde{u}_i] = W_i[\bar{u}/\tilde{u}_i] \quad \text{where } u_j \in \Psi[\bar{H}_j] \quad \forall\, j \neq i$$

and $W_i\,[u/u_i^*] = W_i[\bar{u}/u_i^*]$

Therefore,

$$W_i[\bar{u}/\tilde{u}_i] > W_i[\bar{u}/u_i^*]$$

which contradicts the results shown in [2] and stated above. ■

FOOTNOTES

1. A special case is where M = L and all the α_i are L x L identity matrices. The constraint would then be $\sum_{i=0}^{N} x_i \leq K$.

2. For the Dantzig-Wolfe Decomposition Algorithm, $X = \{x \mid x_i \geq 0, i = 0,1,\ldots,N\}$; for Malinvaud's procedure X is the set of feasible final consumption vectors from the central planner's point of view.

3. The general form of the problem in (2) can accomodate public goods as well. If there were only two divisions, the problem might look like:

$$\max_{x \in X} \pi_0(x_0) + \pi_1(x_1) + \pi_2(x_2)$$

$$\text{s.t.} \quad \begin{aligned} x_0 - x_1 &\leq 0 \\ x_1 - x_0 &\leq 0 \\ x_0 - x_2 &\leq 0 \\ x_2 - x_0 &\leq 0 \end{aligned}$$

so that $K = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and $\alpha_0 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$, $\alpha_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ and $\alpha_2 = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}$

4. The messages m_i^{∞} will be consistent with a function if whenever $m_{it}^{1} = m_{is}^{1}$ for $t \neq s$ (i.e., same resource usage), $m_{it}^{2} = m_{is}^{2}$ (same profit contribution). Clearly, this is easily verifiable and thus the condition is operational for the Center.

5. Let $f: \mathbb{R}^N \to \mathbb{R}$ and $g: \mathbb{R}^N \to \mathbb{R}^L$. The mathematical programming problem

$$\max_{x \in \mathbb{R}^N} f(x)$$

$$\text{s.t. } g(x) \leq 0$$

is decomposible if the Lagrangean function $L(x,y) = f(x) - yg(x)$ (where $y \in \mathbb{R}^L$) can be written as

$$L(x,y) = \sum_{i=1}^{N} [f_i(x_i) - yg_i(x_i)]$$

6. $u \backslash u_i \equiv (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_N)$

$u/\tilde{u}_i \equiv (u_1, \ldots, u_{i-1}, \tilde{u}_i, u_{i+1}, \ldots, u_N)$

7. Clearly, truth-telling is desirable behavior, since it can be shown that following the rules of the price decomposition algorithm leads to a solution of the overall problem as defined in (2).

REFERENCES

1. S.I. Cohen, Incentives and the Choice of Optimal Plans, American Economic Review, Vol. 70, No. 4, September 1980.

2. S.I. Cohen, Incentives, Iterative Communication and Organizational Control, Journal of Economic Theory, Vol. 22, No. 1, February 1980.

3. G. Dantzig and P. Wolfe, The Decomposition Algorithm for Linear Programs, Econometrica 29, No. 4 (1961), p. 767-778.

4. T. Groves, Information, Incentives and the Internalization of Production Externalities, in "Theory and Measurement of Economic Externalities" (S. Lin, ed.), Academic Press, New York, 1975.

5. T. Groves and J. Ledyard, Optimal Allocation of Public Goods: A Solution to the Free-Rider Problem, Econometrica, Vol. 41, No. 4, 1973.

6. E. Malinvaud, Decentralized Procedures for Planning, in "Activity Analysis in the Theory of Growth and Planning" (E. Malinvaud and M.O.L. Bacarach, Eds.), pp. 170-208, Macmillan, London, 1967.

List of Symbols

Script	Greek	Other
$\mathcal{M}$	π	$\mathbb{R}$
	Σ	∞
	τ	ε
	ψ	$\geq$
	α	$\leq$
	λ	$\rangle$
		$\langle$
		$\rightarrow$
		$\sim$
		$\backslash$
		$\wedge$
		$\forall$
		$'$
		$\blacksquare$
		$\overline{\sqcup}$
		$\equiv$
		$\times$

All subscripts "0" are "zero"

Accepted for publication, Management Science

Incentive Properties of Iterative

Decomposition Algorithms

Susan I. Cohen*,**

* Assistant Professor, College of Management, Georgia Institute of Technology, Atlanta, Georgia 30332.

** An earlier version of this paper (with T. Groves) was prepared for presentation at the IEEE Conference on Control and Adaptive processes, Clearwater Beach, Florida, December 1-3, 1976.

ABSTRACT

The Dantzig-Wolfe decomposition algorithm (DWDA) is viewed as a resource allocation procedure for a multi-divisional firm. An evaluation procedure is proposed that provides incentives in equilibrium for the divisional managers to follow the rules of the DWDA even though they are able to "cheat" in a manner that could not be detected. The evaluation procedure for the DWDA is a specific example of a general procedure for the incentive compatible control of large divisionalized organizations using iterative decomposition algorithms.

I. Statement of the Problem

One of the most frequently encountered linear programming problems is the problem solved by the Dantzig-Wolfe decomposition algorithm [3]

$$\max_{x_i} \sum_{i=1}^{N} c_i x_i$$

$$\text{s.t.} \quad A_i x_i \le B_i \qquad i=1,\ldots,N$$

$$\sum_{i=1}^{N} \alpha_i x_i \le K \tag{1.1}$$

$$x_i \ge 0 \qquad i=1,\ldots,N$$

where x_i, c_i and K are vectors, A_i and α_i are matrices, and B_i are scalars.

The Dantzig-Wolfe decomposition algorithm (DWDA) is a "price" decomposition algorithm. A series of subproblems are solved using prices generated by a series of master problems. The revised prices are the dual variables associated with the linking constraints $\sum \alpha_i x_i \le K$. A natural (and frequent) interpretation of (1.1) and the DWDA is that of a divisionalized firm, in which there are externalities, trying to maximize profit using internally generated prices. Division i's "output" is x_i, its contribution to profit is $c_i x_i$ and its local constraint is $A_i x_i \le B_i$. The externalities are created by the linking constraints, where each component of K can be viewed as representing an amount of a scarce resource (e.g. cash, managerial time) that must be allocated to each of these divisions.

Consider a multi-divisional firm that wishes to use the DWDA to solve its resource allocation problem (as modelled by 1.1). The corporate center (hereafter referred to as the Center) sends the divisions "prices" for the scarce resources, and instructs them to choose an output vector that maximizes divisional profit contribution minus the "cost" of the scarce resources.

The Center instructs the divisions to inform it of their profit contribution and resource use at the optimal output vector. This information is used, along with previously obtained information, to calculate revised prices. The process continues until the Center has determined (via an optimality criteria) that further rounds of communication would not lead to a better allocation of the scarce resources, in terms of its overall objective of solving (1.1).

Given the above scenario, the following question is raised: will the divisions find it in their best interests to follow the Center's directives? If the answer is no, then the DWDA is not very useful as a resource allocation procedure. The purpose of this work is the definition of a reward mechanism for the divisions that makes following the Center's directives (i.e., following the rules of the DWDA) in the best interests of the divisions.

There is a growing literature in economics that deals with the provision of incentives in various contexts (see Cohen [2] for a discussion). Arrow [1] defined the problem of organizational control as consisting of two parts: (1) the choice of <u>operating (behavioral) rules</u> for the members of the organization, and (2) the choice of <u>enforcement rules</u> that induce the members of the organization via the division's evaluation to follow the operating rules. Operating rules are made up of information exchange (communication) and decision rules. Enforcement rules must depend on the divisions' actions and be such that a division maximizes its evaluation when it follows the operating rules. In most of the previous work on incentives, the operating rules have been such that communication takes place on a one-time only basis, and the usefulness of the specified enforcement rules relies on that property.

If the DWDA is viewed as a specification of operating rules, iterative communication may require the definition of different enforcement rules. If the same type of rules are used as in the non-iterative case, the results obtained may be different. In devising appropriate enforcement rules for the DWDA, we have relied on the general theory of organizational control developed by Groves [4,5] where communication is non-iterative. The extension to very general programming models and to iterative decomposition algorithms is due to Cohen [2]. In the current work, as in [2], we obtain weaker results than Groves because, when communication is iterative rather than one-time only, the Center has much less information available when it makes its final decision.

Since the results we obtain are weaker than in the non-iterative case, why not just redefine the DWDA so that communication takes place all at once? In reality, the divisions of a firm probably cannot compute and/or store all at once the amount of information this would require. A division may be able to tell the Center how much of a scarce resource it wants at a particular price, but to expect it to be able to do so for every possible price is unreasonable. In addition, since presumably the cost of calculation is not zero, why should the division do a large amount of what turns out to be unnecessary calculation? Most real world planning procedures have iterative communication [6] and thus we must be able to provide incentives in that context as well as in the non-iterative case.

Jennergren [7] has investigated the provision of incentives for the DWDA. He demonstrated that a particular reward scheme for division managers does not work. The division managers may have an incentive to "cheat" (not follow the rules) and could do so in an undetectable manner. In Section II, we define a class of evaluation measures under which there is no incentive to cheat; in Section III, we view the divisions' problem in terms of an

N-person game and show that following the rules is a Nash equilbrium that dominates in the class of Nash equilibria. In Section IV, we discuss the problem of terminating the DWDA before optimality is reached. Concluding remarks are contained in Section V.

II. Incentive Compatible Divisional Evaluations

For ease of exposition, we define the following functions:

$$\pi_i(y_i) \equiv \{\max_{x_i} c_i x_i \text{ s.t. } A_i x_i \leq B_i,\ \alpha_i x_i \leq y_i,\ x_i \geq 0\} \tag{2.1}$$

It is straightforward that $\pi_i(\cdot)$ is a concave, polyhedral and continuous function of its argument (Cohen [2]). This definition allows us to rewrite (1.1) as

$$\max_{y_i} \sum_{i=1}^{N} \pi_i(y_i) \quad \text{s.t.} \sum_{i=1}^{N} y_i \leq K \tag{2.2}$$

Let p^1 be the initial price vector in the DWDA. If the divisions were following the rules they would then solve

$$\max_{y_i} \ \pi_i(y_i) - p^1 y_i \tag{2.3}$$

and send y_i^1 and $\pi_i(y_i^1)$ to the Center (where y_i^1 solves (2.3)). At any iteration t, the division would solve (2.3) with p^1 replaced by p^t and send y_i^t and $\pi_i(y_i^t)$ to the Center. The Center would calculate revised prices p^{t+1} by solving the following,

$$\begin{aligned} \max_{\theta_{i\tau}} \sum_{i=1}^{N} \sum_{\tau=1}^{t} \ \theta_{i\tau}\, \pi_i(y_i^\tau) & \\ \text{s.t.} \quad \sum_{i=1}^{N} \sum_{\tau=1}^{t} \theta_{i\tau}\, y_i^\tau \leq K & \\ \sum_{\tau=0}^{t} \theta_{i\tau} = 1 \qquad & i=1,\ldots,N \\ \theta_{i\tau} \geq 0 \qquad & i=1,\ldots,N \\ & \tau=0,1,\ldots,t \end{aligned} \tag{2.4}$$

where $\langle\{\theta_{i\tau}^{t+1}\}\rangle_{i=1}^{N}$ is the solution at iteration t, and p^{t+1} is the vector of dual variables associated with the first constraint in (2.4).[1]

The Center's decision consists of an allocation y_i^* of scarce resources to each of the divisions[2]

$$y_i^* = \lim_{t\to\infty} \sum_{\tau=1}^{t} \theta_{i\tau}^{t+1} \cdot \pi_i(y_i^\tau) \tag{2.5}$$

Suppose now that the divisions don't follow the rules of the DWDA. At each iteration they send what appears to be the solution to (2.3) (where p^1 is replaced by p^t), but is not. Let a_{it} be division i's "resource use" at t and b_{it} be division i's "profit contribution" at t. The Center then solves (2.4) where y_i^τ is replaced by $a_{i\tau}$ and $\pi_i(y_i^\tau)$ is replaced by $b_{i\tau}$ for all $\tau \le t$, i.e. the Center believes the divisions to be following the rules of the DWDA and behaves accordingly. Clearly, the revised price vector p^{t+1}, the optimal weights $\theta_{i\tau}^{t+1}$ as well as the final allocation of resources y_i^* are now functions of the divisions' reported resource use and profit contribution. Define $a_i^t \equiv (a_{i1},\ldots,a_{it})$, $a^t \equiv (a_1^t,\ldots,a_N^t)$, $b_i^t \equiv (b_{i1},\ldots,b_{it})$, and $b^t \equiv (b_1^t,\ldots b_N^t)$. Then

$$\begin{aligned} p^{t+1} &\equiv p^{t+1}(a^t,b^t) \\ \theta_{i\tau}^{t+1} &\equiv \theta_{i\tau}^{t+1}(a^t,b^t) \\ y_i^* &\equiv y_i^*(a^t,b^t) = \lim_{t\to\infty} \sum_{\tau=1}^{t} \theta_{i\tau}^{t+1}(a^t,b^t) \cdot a_{i\tau} \end{aligned} \tag{2.6}$$

We can now define <u>evaluation measures</u> $R_i(\cdot)$ for the divisions, where each division's evaluation depends on its realized profit contribution and the reported profit contribution of the other divisions. It should be noted that these evaluations do not necessarily represent real resource flows. One way to view then is as a scoring system for division managers, where the "score" is monotonically transformed into the arguments of the division manager's utility function. If they do represent real resource flows, then one must be concerned with the question of a "balanced budget". A straightforward transformation of the evaluation measure defined below is sufficient to guarantee a balanced budget in all situations of interest. The following definition is made for ease of exposition.

$$S_i^t \equiv \{(a_i^t, b_i^t) \;\Big|\; \sum_{\tau=1}^{t} \theta_\tau b_{i\tau} \leq b_{is} \text{ for all } \theta_\tau \geq 0 \text{ such that } \sum_{\tau=1}^{t} \theta_\tau a_{i\tau} = a_{is} \text{ and } \sum_{\tau=1}^{t} \theta_\tau = 1, \text{ for all } s=1,\ldots,\tau\} \tag{2.7}$$

The set S_i^t contains all possible "resource allocations" and "profit contributions" that could be consistent with division i following the rules of the DWDA. If π_i is division i's realized profit contribution, then

$$R_i(\pi_i, (a^\infty, b^\infty)) \equiv \begin{cases} \pi_i + \sum_{j\neq i} \lim_{t\to\infty} \sum_{\tau=1}^{t} \theta_{j\tau}^{t+1}(a^t, b^t) \cdot b_{j\tau} \\ \quad \text{if } (a_i^t, b_i^t) \in S_i^t \text{ for all } t \\ \quad \underline{\text{and}} \\ \quad \pi_i \geq \lim_{t\to\infty} \sum_{\tau=1}^{t} \theta_{i\tau}^{t+1}(a^t, b^t) \cdot b_{i\tau} \\ -\infty \text{ otherwise} \end{cases} \tag{2.8}$$

The Center must believe that division i is following the rules of the DWDA and division i's realized profit contribution must be at least as large as its reported contribution.[3,4] The $R_i(\cdot)$ appear to be a type of "profit sharing" since each division's evaluation is the sum of its actual profit contribution and the other divisions' reported profit contribution. The difference between traditional profit sharing and the evaluations defined in (2.8) is that with these new evaluations, each division's maximizing behavior does not depend on the assumption of rationality of the other divisions. Division i's maximizing behavior is thus independent of whether or not the other N-1 divisions are maximizing their own evaluations.

Given the evaluations defined in (2.8), the divisions must now decide how to respond to the revised prices sent by the Center at each iteration. The division wants to ensure that the Center's final allocation of resources maximizes the division's evaluation. The division chooses, for each t, a <u>response rule</u> or function u_{it} that determines at t how division i should respond to the price vectors it has received from the Center. If each p^t has L components (i.e., there are L scarce resources), then

$$u_{it} : \mathbb{R}^{L \cdot t} \to \mathbb{R}^{L+1} \tag{2.9}$$

A <u>strategy</u> u_i is an infinite sequence of response rules, one for each possible t.

$$u_i \equiv \{u_{it}\}_{t=1}^{\infty} \tag{2.10}$$

The strategies for each division must be chosen from the following set U:

$$U \equiv \{\text{the space of infinite sequences defined by (2.9) and (2.10)}\} \tag{2.11}$$

Each a_{it} and b_{it} can be expressed as a function of the response rules of all the divisions at previous iterations, the response rule of division i at t, the rule for choosing revised prices and the initial price vector p^1.

Therefore

$$
\begin{aligned}
a_{it} &: U^N \to \mathbb{R}^L \\
b_{it} &: U^N \to \mathbb{R} \quad .
\end{aligned}
\tag{2.12}
$$

Letting $u \equiv (u_1,\ldots,u_N)$, we can define $a_i^t(u)$, $a^t(u)$, $b_i^t(u)$, and $b^t(u)$ analagously to a_i^t , a^t, b_i^t and b^t where a_{it} and b_{it} are replaced by $a_{it}(u)$ and $b_{it}(u)$ in the former; in addition $a^{\infty}(u)$ and $b^{\infty}(u)$ are the infinite dimensional strategy-dependent vectors of resource use and profit contribution, respectively, for all divisions at all iterations.

III. The Divisions' Problem as an N-person Game

The choice of divisional strategies can be thought of in game theoretic terms. In order to do so we must define payoff functions $W_i[\cdot]$ for each division.

$$W_i[u] \equiv R_i(\pi_i(y_i^* (a^\infty(u),\ b^\infty(u)))\ ,\ (a^\infty(u),\ b^\infty(u))) \tag{3.1}$$

These payoff functions and the set of strategies U define an N-person (non-zero sum) game in normal form, which we assume is played non-cooperatively.

The following Lemma is given without proof, which follows immediately from the convergence properties of the DWDA. (The proof of all subsequent Theorems appear in the Appendix.)

LEMMA 1 If u_i^* , for all i, chooses (a_{it},b_{it}) according to the rules of the DWDA for all t, then $(x_1^*(u^*),\ldots,x_N^*(u^*))$ solves (1.1), where $c_i x_i^* (u) = \pi_i(y_i^* (u))$.

A dominant strategy for a player in a non-cooperative game is a strategy that maximizes the player's payoff no matter what strategies that the other players use, i.e., $\tilde{u}_i$ is a dominant strategy if $W_i[u/\tilde{u}_i] \geq W_i[u]$ for all $u_i \in U$, for any $u \backslash u_i \in U^{N-1}$.[5] Ideally, one would like to show the existence of a dominant strategy equilibrium for a game. In the previously cited work of Groves, where communication is non-iterative, these type of equilibria do indeed exist. The loss of information due to iterative communication prevents the existence of such equilibria in the present game. Another solution concept for a non-cooperative game is the Nash equilibrium. The N-tuple u^* is a Nash equilibrium if $W_i[u^*] \geq W_i[u^*/u_i]$ for all $u_i \in U$, for all i. We can show for the current game that following the rules of the DWDA is a Nash equilibrium that dominates all other Nash equilibria for every division i. Any such dominant Nash equilibrium is a likely outcome of the game; we show that they lead to decisions that solve the overall firm problem.

<u>THEOREM 1</u> If u_i^*, for all i, chooses (a_{it}, b_{it}) according to the rules of the DWDA, then

a) u^* is a Nash equilibrium

b) if $\hat{u}$ is any other Nash equilibrium the payoff for every division is at least as great under u^* as it is under $\hat{u}$.

<u>THEOREM 2</u> If $\hat{u}$ is a Nash equilibrium such that the payoff under $\hat{u}$ is at least as great as under any other Nash equilibrium u for every division, then $(x_1^* (\hat{u}), \ldots, x_N^* (\hat{u}))$ solves (1.1) where $x_i^* (\hat{u})$ is defined as in Lemma 1.

Therefore, in Theorem 1, we have shown that it is possible to provide incentives for the divisions to follow the rules of the DWDA and in Theorem 2, we show that such "rational" behavior on the part of the divisions leads to a solution of the overall firm problem (1.1).

IV. Terminating the DWDA

The results in Section III depend upon the DWDA being run until the optimality criteria for the algorithm are satisfied. If the algorithm is terminated before convergence, the results we obtain are no longer necessarily valid in general. The divisions then face a problem of decision making under uncertainty, where their respective probability estimates about when the algorithm (i.e., communication) will be terminated determine their optimal strategies. Following the rules of the DWDA is no longer necessarily in the best interests of the divisions, since end-gaming may now be optimal. The solution of the problem of termination before optimality is beyond the scope of the present work.

V. Summary

We have shown that when the DWDA is viewed as a decentralized resource allocation procedure, it is possible to provide incentives for divisional managers to follow the rules of the DWDA and that "rational" behavior on the part of the divisions leads to decisions that solve the overall problem.

However, the results obtained here are "weaker" than those obtained by Groves [4] when communication was non-iterative. The difference is due to the different information possessed by the Center when it makes its decisions. In the earlier work, the Center knows each divisions' strategy in its entirety. In the current work, the Center has only a partial approximation of the strategy. The reasonableness of the results presented here depend upon the acceptability of the Nash equilibrium as an outcome concept. This, we maintain, is primarily an empirical question.

Appendix

Proof of Theorem 1

As a preliminary to proving (a) and (b), we must show that

$$\pi_i(y_i^*(a^\infty(u/u_i^*),\ b^\infty(u/u_i^*))) = \lim_{t\to\infty} \sum_{\tau=1}^{t} \theta_{i\tau}^{t+1}(a^t(u/u_i^*),\ b^t(u/u_i^*)) \cdot b_{i\tau}(u/u_i^*) \tag{A.1}$$

Since $\pi_i(\cdot)$ is concave and polyhedral, there exists a finite T such that

$$\pi_i(\sum_{\tau=1}^{T} \theta_{i\tau}^{T+1}(a^T(u/u_i^*),\ b^T(u/u_i^*)) \cdot a_{i\tau}(u/u_i^*)) = \sum_{\tau=1}^{T} \theta_{i\tau}^{T+1}(a^T(u/u_i^*),\ b^T(u/u_i^*)) \cdot b_{i\tau}(u/u_i^*)$$

Taking limits and using continuity of $\pi_i(\cdot)$ and the definition of $y_i^*(a^\infty, b^\infty)$ we have

$$\lim_{T\to\infty} \sum_{\tau=1}^{T} \theta_{i\tau}^{T+1}(a^T(u/u_i^*),\ b^T(u/u_i^*)) \cdot b_{i\tau}(u/u_i^*) =$$

$$\lim_{T\to\infty} \pi_i(\sum_{\tau=1}^{T} \theta_{i\tau}^{T+1}(a^T(u/u_i^*),\ b^T(u/u_i^*)) \cdot a_{i\tau}(u/u_i^*)) =$$

$$\pi_i(\lim_{T\to\infty} \sum_{\tau=1}^{T} \theta_{i\tau}^{T+1}(a^T(u/u_i^*),\ b^T(u/u_i^*)) \cdot a_{i\tau}(u/u_i^*)) =$$

$$\pi_i(y_i^*(a^\infty(u/u_i^*),\ b^\infty(u/u_i^*)))$$

which is the desired result.

(a) u^* is a Nash equilibrium if for all i, $W_i[u^*] \geq W_i[u^*/u_i]$ for all u ε U. From (A.1) and the definition of $W_i[u]$, we have that for all i

$$W_i[u^*] = \sum_{j=1}^{N} \pi_j(y_j^*(a^\infty(u^*), b^\infty(u^*)))$$

From Lemma 1,

$$W_i[u^*] \geq \sum_{j=1}^{N} \pi_j(y_j) \quad \text{for all } y_j \text{ such that} \quad \sum_{j=1}^{N} y_j \leq K \tag{A.2}$$

In particular,

$$W_i[u^*] \geq \sum_{j=1}^{N} \pi_j(y_j^*(a^\infty(u^*/u_i), b^\infty(u^*/u_i))) \quad \text{for all } u_i \in U$$

Using (A.1) for all j≠i, we get the desired result,

$$W_i[u^*] \geq \pi_i(y_i^*(a^\infty(u^*/u_i), b^\infty(u^*/u_i))) +$$

$$\sum_{j\neq i} \lim_{t\to\infty} \sum_{\tau=1}^{t} \theta_{j\tau}^{t+1}(a^t(u^*/u_i), b^t(u^*/u_i)) \cdot b_{j\tau}(u^*/u_i) \geq$$

$$W_i[u^*/u_i] \quad \text{for all } u_i \in U.$$

(b) From (A.2),

$$W_i[u^*] \geq \sum_{j=1}^{N} \pi_j(y_j^*(a^\infty(\hat{u}),\ b^\infty(\hat{u})))$$

for any Nash equilibrium $\hat{u}$.

For any Nash point $\hat{u}$, it must be true that for all j

$$\pi_j(y_j^*(a^\infty(\hat{u}),\ b^\infty(\hat{u}))) \geq \lim_{t\to\infty} \theta_{j\tau}^{t+1}(a^t(\hat{u}),\ b^t(\hat{u})) \cdot b_{j\tau}(\hat{u})$$

Therefore, for all i,

$$W_i[u^*] \geq \pi_i(y_i^*(a^\infty(\hat{u}),\ b^\infty(\hat{u}))) + \sum_{j\neq i} \lim_{t\to\infty} \theta_{j\tau}^{t+1}(a^t(\hat{u}),\ b^t(\hat{u})) \cdot b_{j\tau}(\hat{u}) = W_i[\hat{u}]. \quad \square$$

Proof of Theorem 2

(by contradiction). Let $\hat{u}$ be a Nash equilibrium whose payoff is at least as great as any other Nash equilibrium u for all divisions. With u^* defined as in Theorem 1 and using the results of that Theorem, we have

$$W_i[\hat{u}] = W_i[u^*] \quad \text{for all } i \tag{A.3}$$

From the proof of Theorem 1

$$\sum_{i=1}^{N} \pi_i(y_i^*(a^\infty(u^*),\ b^\infty(u^*))) \geq \sum_{i=1}^{N} \pi_i(y_i^*(a^\infty(\hat{u}),\ b^\infty(\hat{u}))) \tag{A.4}$$

Assume $(x_1^*(\hat{u}),\ldots,x_N^*(\hat{u}))$ does not solve (1.1). Then strict inequality holds in (A.4). For every i, $W_i[u^*]$ equals the left-hand side of (A.4). Therefore, using (A.3), $W_i[\hat{u}]$ equals the left hand side of (A.4) for all i. Using the definition of $W_i[\hat{u}]$ and summing over all i, we get

$$\sum_{i=1}^{N} [\pi_i(y_i^*(a^\infty(\hat{u}),\ b^\infty(\hat{u}))) + \sum_{j\neq i} \lim_{t\to\infty} \sum_{\tau=1}^{t} \theta_{j\tau}^{t+1}(a^t(\hat{u}),\ b^t(\hat{u})) \cdot b_{j\tau}(\hat{u})] >$$

$$N\sum_{i=1}^{N} \pi_i(y_i^*(a^\infty(\hat{u}),\ b^\infty(\hat{u}))) \tag{A.5}$$

Since $\hat{u}$ is a Nash equilibrium, for all i

$$\pi_i(y_i^*(a^\infty(\hat{u}),\ b^\infty(\hat{u}))) \geq \lim_{t\to\infty} \sum_{\tau=1}^{t} \theta_{i\tau}^{t+1}(a^t(\hat{u}),\ b^t(\hat{u})) \cdot b_{i\tau}(\hat{u}) \tag{A.6}$$

Using (A.5) and (A.6) we get

$$\sum_{i=1}^{N} \pi_i(y_i^*(a^\infty(\hat{u}),\ b^\infty(\hat{u}))) > \sum_{i=1}^{N} \pi_i(y_i^*(a^\infty(\hat{u}),\ b^\infty(\hat{u})))$$

which is our contradiction. Therefore $(x_1^*(\hat{u}),\ldots,x_N^*(\hat{u}))$ solves (1.1). □

Footnotes

1. The indices in the last two sets of constraints in (2.4) go from 0 to t instead of 1 to t in order to insure a feasible solution.

2. The limits in (2.5) as well as all subsequent limits exist by the definition of the $\theta_{i\tau}$, i.e., the sum is bounded by K for all t.

3. a^{∞} and b^{∞} are the infinite dimensional vectors of all the divisions resource uses and profit contributions, respectively, for all iterations.

4. If $R_i(\pi_i,a,b)$ is modified to $R_i'(\pi_i,a,b) \equiv \lambda_i R_i(\pi_i,a,b)$ where $\sum_{i=1} \lambda_i = 1$ and the λ_i are preditermined, budget balance is no longer a problem. [See Cohen [2]].

5. $u/\tilde{u}_i \equiv (u_1,\ldots,u_{i-1},\tilde{u}_i,u_{i+1},\ldots,u_N)$
   $u\backslash u_i \equiv (u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_N)$

References

1. Arrow, K., "Control in Large Organizations," Management Science, Vol. 10, No. 3 (1964), pp. 397-408.

2. Cohen, S., "Incentive Compatible Control of the Multidivisional Firm with Iterative Communication," Ph.D. Dissertation, Northwestern University (1977).

3. Dantzig, G. and P. Wolfe, "The Decomposition Algorithm for Linear Programming," Econometrica, Vol. 29, No. 4 (Oct., 1961), pp. 767-778.

4. Groves, T., "Incentive Compatible Control of Decentralized Organizations," in Y.C. Ho and S. Mitter (eds.), Directions in Large-Scale Systems, Plenum (1976), pp. 149-185.

5. Groves, T., "Incentives in Teams," Econometrica, Vol. 41, No. 4 (July 1973), pp. 617-631.

6. Heal, G.M., The Theory of Economic Planning, North-Holland, (1973).

7. Jennergren, L., "Studies in the Mathematical Theory of Decentralized Resource-Allocation," Ph.D. Dissertation, Stanford University (1971).